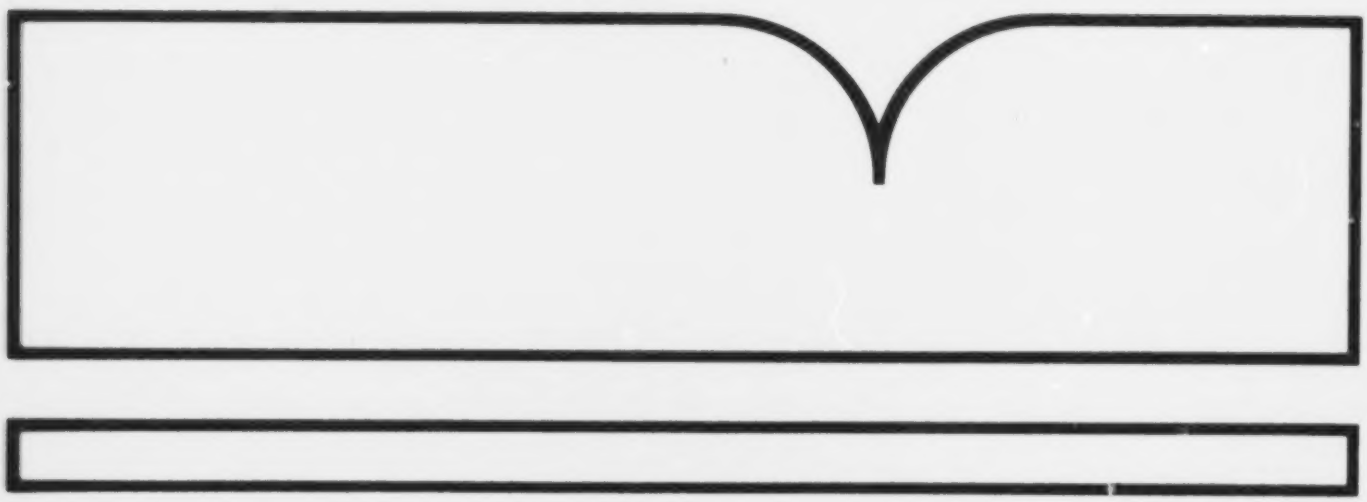
PB87-178349

Test Method Equivalency Petitions
A Guidance Manual

(U.S.) Environmental Protection Agency
Washington, DC

Feb 87



U.S. Department of Commerce
National Technical Information Service
NTIS

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. EPA/530-SW-87-008	2.	3. Recipient's Accession No. PB87 178349/AS
4. Title and Subtitle Test Method Equivalency Petitions a Guidance Manual			5. Report Date February 1987 6.
7. Author(s) Florence M. Richardson			8. Performing Organization Rept. No.
9. Performing Organization Name and Address EPA 401 M Street, S.W. Washington, D.C. 20460			10. Project/Task/Work Unit No. 11. Contract(C) or Grant(G) No. (C) (G)
12. Sponsoring Organization Name and Address EPA			13. Type of Report & Period Covered 14.

15. Supplementary Notes

16. Abstract (Limit: 200 words)

This manual provides guidance to parties who wish to submit a test method petition. It explains in detail, the information a test method equivalency must include. By following the guidance in this manual, a petitioner should be able to develop a petition that satisfies the regulations and that EPA can review expeditiously.

17. Document Analysis a. Descriptors

b. Identifiers/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Unclassified	Unclassified	117
	20. Security Class (This Page) Unclassified	22. Price

(See ANSI-Z39.18) See Instructions on Reverse

OPTIONAL FORM 272 (4-77)
(Formerly NTIS-35)
Department of Commerce

United States
Environmental Protection
Agency

Office of
Solid Waste
Washington, DC 20460

EPA/530-SW-87-008
OSWER Policy Directive
No. 9433.00-2
February 1987

Solid Waste

PB87-178349

EPA

Test Method Equivalency Petitions

A Guidance Manual

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL
INFORMATION SERVICE
SPRINGFIELD, VA 22161

TEST METHOD EQUIVALENCY PETITIONS:

A GUIDANCE MANUAL

OFFICE OF SOLID WASTE
U.S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

1986

CONTENTS

1. Introduction 1

2. Regulatory Requirements 2

3. The Petition Process and EPA Evaluation Criteria 3

4. Statistical Procedures for Evaluating Test Method Equivalency . 7

5. How to Prepare a Petition. 11

6. Checklist of Petition Requirements 21

References 25

Appendices

A. 40 CFR 260, Subpart C--Rulemaking Petitions A-1

B. Statistical Procedures to Evaluate Test Method Equivalency . . . B-1

C. Section Ten--Quality Control/Quality Assurance C-1

D. Example Test Method Equivalency Petition D-1

SECTION 1

INTRODUCTION

The EPA manual, "Test Methods for Evaluating Solid Waste" (SW-846) provides an up-to-date, unified source of information and methods on sampling, analysis, and quality assurance for compliance with the Resource Conservation and Recovery Act (RCRA) regulations. It is a collection of sampling and analysis methods and procedures approved for evaluating the properties of wastes and environmental media, and for monitoring the efficacy of treatment. SW-846 is incorporated by reference in the RCRA regulations.

The RCRA regulations (40 CFR 260.20) establish procedures by which persons may petition EPA to approve the use of alternative or equivalent testing procedures when conducting testing under RCRA. Throughout this guidance manual, the term "test method equivalency petition" is used to denote all petitions that propose methods to replace or supplement any of the test methods described in SW-846.

Any person or organization may submit a petition to request approval for a proposed test method. Petitioning entails the submission to EPA of comprehensive information describing the method, data from tests designed to evaluate equivalency with existing methods, and a statistical analysis of the equivalency test data. EPA evaluates test method equivalency petitions for completeness, applicability, and technical quality. In addition, the Agency reserves the right to conduct an independent statistical analysis of the equivalency data. Methods that pass the evaluation and do not receive substantial negative public comment will be approved, and may be published in SW-846 for public use (e.g., if it can be applied to multiple sites).

The purpose of this manual is to provide guidance to parties who wish to submit a test method petition. It explains, in detail, the information a test method equivalency petition must include. The following chapters and appendices provide:

- An explanation of the RCRA regulations that require a test method equivalency petition
- A discussion of the process and procedures by which a petition is submitted to and reviewed by EPA
- A description of basic statistical procedures to be used for evaluating test method equivalency
- A description of how to prepare a petition, with a suggested petition format
- A checklist to help ensure the completeness of the petition
- A discussion of proper experimental designs and statistical analyses for more complex petitions
- An example of a test method equivalency petition.

By following the guidance in this manual, a petitioner should be able to develop a petition that satisfies the regulations and that EPA can review expeditiously.

SECTION 2

REGULATORY REQUIREMENTS

Title 40 of the Code of Federal Regulations (CFR), Part 260, Subpart C, defines the procedures and information required for rulemaking petitions. In particular, Sections 260.20 and 260.21 specify information that is required for test method equivalency petitions.[1]

Section 260.20 of the regulation contains information that applies to all types of petitions including test method equivalency petitions. It describes general information the petition must include and outlines the decisionmaking procedures EPA follows to approve or deny petitions. Section 260.21 establishes specific information needs for test method equivalency petitions; i.e., what data and information EPA must have to determine if a proposed test method is equal or superior to a corresponding SW-846 test method.

By providing the necessary information, the petitioner can meet the objectives of the regulations, which are:

- To ensure that appropriate, accurate, and precise test methods are used
- To ensure the comparability of all hazardous waste test data gathered in support of EPA's regulatory program.

---

[1]Appendix A contains the text of 40 CFR 260.20 and 260.21.

SECTION 3

THE PETITION PROCESS AND EPA EVALUATION CRITERIA

3.1 DESCRIPTION OF THE PETITIONING PROCESS

This section will acquaint the petitioner with the complete process by the which a test method equivalency petition is submitted to and reviewed by EPA. The discussion is organized around a flowchart (Figure 3-1) that illustrates the sequential steps in the process. The time periods shown in this figure are approximate; they provide only an estimate of the time required for each step in the petition process.

3.1.1 Pre-petition Contact with the EPA

Prior to submitting a petition, the petitioner is advised to contact the EPA Methods Section staff within the Office of Solid Waste (OSW) at (202) 382-4761. Such contact can be used to identify whether the proposed method is actually different from an existing test method. For example, a modification of measurement technique or a modification to the equipment may already be within the scope of a current method and may not constitute a method variance requiring a petition. Alternatively, the use of different measurement techniques or new types of equipment frequently requires petitioning for approval. Familiarity with the SW-846 document and a phone call or correspondence to the EPA Methods Section staff can normally resolve the question of how much the proposed method varies from the approved method.

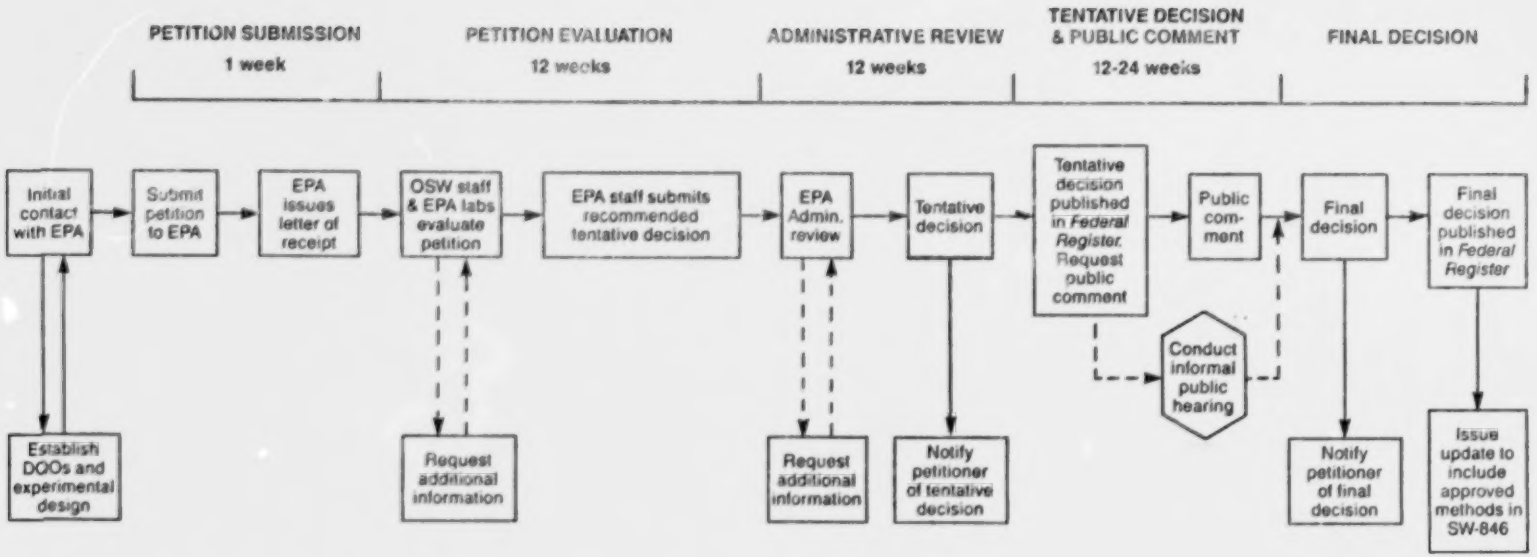
Unless there is a great deal of existing data that demonstrate the equivalency or capabilities of a proposed test method, a specific program of equivalency testing must be undertaken by the petitioner. During this stage of pre-petition contact, the petitioner is urged to develop an experimental design that will satisfy the data quality objectives (DQOs) specified by the EPA Program Office. (See Section 4.)

3.1.2 Submission of the Petition

Any applicant may request approval for an alternative test method. Requests should be submitted in triplicate and by certified mail to: Chief, Methods Section, Office of Solid Waste (WH-562B), U.S. EPA, 401 M Street, SW, Washington, DC 20460. The EPA Methods Section Chief will then acknowledge receipt of the petition by sending a letter to the petitioner.

3.1.3 EPA Evaluation and Preliminary Recommendations

The initial technical review of petitions is conducted by EPA staff in OSW and several EPA laboratories. This review does not include actual laboratory testing but is limited to a critical analysis of the reported test results and the associated statistical analyses. If sufficient information or data are not available for an acceptable review, EPA returns the petition to the applicant with specific requests for additional information. Assuming no



NOTE: Dashed lines indicate optional procedures.

Figure 3-1. Petition procedures flowchart.

OSWER Policy Directive #9433.00-2

additional testing and data submission are required of the applicant, the EPA technical evaluation should take about 12 weeks.

Following technical review of the petition, the EPA Methods Section Chief recommends approval or denial of the proposed method through a memorandum. This memorandum contains a complete review of the decisionmaking process including a technical rationale for the proposed action.

3.1.4 EPA Administrative Review

The recommendations of the OSW staff are reviewed by the methods workgroup and then circulated to other EPA Program Offices (e.g., Office of Water, Office of Toxic Substances) for compliance with national policies. If more information is needed, the Agency may request additional information from the petitioner. The Agency then makes a tentative decision to grant or deny the petition, notifies the petitioner accordingly, and prepares a notice of the tentative decision for publication in the Federal Register.

3.1.5 Notice of Tentative Decision and Public Comment

The notice of the tentative decision published in the Federal Register is accompanied by a request for public comment. Comments may be submitted in writing, or commenters may request a public hearing. Regulations governing EPA decisionmaking on petitions state that "the Administrator may, at his discretion, hold an informal public hearing to consider oral comments on the tentative decision" (40 CFR Part 260.20(d)). Although an informal public hearing must satisfy all fairness criteria in public participation, it differs from a formal public hearing in the following respects:

- Legal standing: a judge does not preside at an informal public hearing.
- Scheduling: informal public hearings usually do not schedule witnesses in advance.
- Transcript: EPA does not prepare a complete written or taped transcript of informal public hearings.

The petitioner is expected to participate in any public hearing regarding the proposed test method.

3.1.6 Publication of Final Decision and Incorporation into SW-846

After evaluating all public comments and EPA staff recommendations, the Agency makes a final decision and publishes (in the Federal Register) a regulatory amendment or denial of the petition. If the proposed test method is approved, and if it is likely to have widespread applications, the method may be incorporated in "Test Methods for Evaluating Solid Waste: Physical/Chemical Methods," SW-846. Updates (procedures for newly approved methods) and revisions (changes in existing methods) to SW-846 are issued as necessary by EPA to all persons or organizations holding subscriptions to SW-846 through the Government Printing Office.

3.2 DESCRIPTION OF EPA'S TECHNICAL EVALUATION CRITERIA

When a petition is submitted to EPA, the Agency must determine whether the proposed test method is at least equivalent to an existing method in that it yields comparable results. In this case, equivalency is defined as "equal to or superior to the corresponding SW-846 method in terms of its sensitivity, accuracy, and precision (i.e., reproducibility)" (40 CFR 260.21). To determine equivalency, EPA evaluates test method equivalency petitions using the following three criteria: (1) applicability of the test method, (2) adequacy of documentation, and (3) statistical comparability of proposed and approved test methods.

Test method applicability refers to the appropriateness of the proposed for test method for the type of waste the method is designed to detect. Thus, example, a method designed to detect a constituent normally found in waste sludge should be applicable to various types of sludge. In addition, the petition should be clear regarding how broad an approval is desired, because the method can only be approved for the universe of wastes to which test data are applicable or validly extrapolatable. Adequacy of documentation refers to the completeness with which the applicant has fulfilled the information and data requirements outlined in Sections 5 and 6 of this guidance manual.

The final criterion, statistical comparability, is used to evaluate equivalency test data in terms of specific data quality objectives such as precision, bias, and method sensitivity. Specific numerical criteria for determining statistical equivalency will be developed in coordination with the EPA Program Office, with each proposed method considered on a case-by-case basis. Section 4 provides more detailed information on the statistical analysis aspects of equivalency petitioning and guides the petitioner through the development of an appropriate, statistically sound testing program.

SECTION 4

STATISTICAL PROCEDURES FOR EVALUATING TEST METHOD EQUIVALENCY

4.1 INTRODUCTION

This section and Appendix B address statistical issues in the petitioning process.

The Agency recognizes that no single specification of data requirements (type and amount) for demonstrating test method equivalency is appropriate for all test methods. As a result, the petitioner has various options regarding information submitted in support of a proposed method. For instance, a test method could, under certain conditions, be approved on the basis of existing data on the proposed method only. Thus, the time and cost of generating special test data could be eliminated. Under other conditions, a controlled experiment involving both the proposed method and the approved method may be necessary to assess the adequacy of a particular method.

4.2 DATA QUALITY OBJECTIVES

In a recently prepared guidelines document for validating measurement methods (2), EPA states that "Specifications of the data quality needed for a particular data collection activity are called data quality objectives (DQOs). DQOs are definitive, quantitative or qualitative statements developed by data users about the accuracy, precision, representativeness, comparability and completeness of measurement data needed to support their specific decisions." (This document (2) and others (3,4) provide guidance in validation and equivalency.)

It is the responsibility of the EPA/OSW program office to specify DQOs. These objectives are not fixed numbers but may vary depending on the type of measurement under consideration. Once DQOs are specified, the petitioner must then select (or work with the OSW staff to develop) an experimental design that will enable the petitioner to collect adequate test data to determine whether the proposed method complies with the DQOs.

While it is the Agency's intent to allow appropriate flexibility in determining DQOs, the Agency expects uniformity in certain aspects of formulating the problem. Key data quality objectives will involve the bias and precision of the proposed test methods. However, frequently, DQOs will also include detection limits or limits of reliable measurement and the significance level and power of hypothesis tests about bias and precision.

A literal interpretation of 40 CFR 260.21 is misleading; the proposed test method does not always have to be equal or superior (i.e., more accurate) to an EPA-approved test method - regardless of application - in order to be approved. Existing approved test methods are sometimes more accurate than needed for specific Agency uses. In these cases, the DQOs should reflect the Agency's measurement needs rather than the capabilities of existing approved methods. Thus, a proposed test method does not necessarily have to be at

least as good as the approved method to be useful (and acceptable) for certain testing purposes.

In developing acceptance criteria based upon DQOs, it is important to keep in mind that, for example, precision can be improved by taking repeated measurements of a quantity and estimating it as the average of the measurements. For instance, the average of four measurements is twice as precise as a single measurement. This is important to remember if the proposed method is somewhat less precise (based on single estimates) than EPA criteria or than offered by the existing approved method. Were the method simple and inexpensive enough that multiple independent measurements could be made without increasing the cost or time of the overall analyses, the resulting precision from average values might be quite acceptable for the desired use. This principle must be applied cautiously since repeat measurements must indeed be truly independent and reflect appropriate sources of variation. Replicates often fail to meet these requirements.

In general, data quality objectives (or acceptance criteria based upon them) can be classified as either absolute or comparative. Absolute DQOs reflect the Agency's intrinsic needs in the given situation and do not involve a direct comparison of the proposed method to some approved test method. On the other hand, comparative DQOs typically state that the proposed test method must be equivalent (or superior) to an approved test method.

The statistical methods used to assess compliance with absolute and comparative DQOs differ. In the absolute case, the precision (σ^2_{new}) and bias (b_{new}) of the proposed test method are compared with numeric criteria (DQOs) for precision and bias (σ^2_0 and b_0) established at the beginning of the petitioning process. This comparison is carried out by testing the null hypothesis (H_1) that the precision offered by the proposed method is no worse than that established as a program requirement and the null hypothesis (H_2) that the bias of the new method is no greater than what has been established as a progran requirement:

$$H_1: \quad \sigma^2_{new} \leq \sigma^2_0$$

$$H_2: \quad |b_{new}| \leq b_0$$

In the comparative case, the new test method's precision and bias are compared with the approved method's precision and bias (σ^2_{old} and b_{old}). This comparison uses the null hypothesis (H_3) that there is no difference between the precision of the proposed method and that demonstrated with the approved method it is proposed to replace. This comparison also uses the null hypothesis (H_4) that there is no difference between the bias encountered in the use of the proposed method than that is expected in use of the approved method:

$$H_3: \quad \sigma^2_{new} = \sigma^2_{old}$$

$$H_4: \quad b_{new} = b_{old}$$

Note that the way the null hypotheses above are formulated, the null hypotheses must be rejected to disapprove the new method. EPA is more concerned about the errors of falsely accepting these hypotheses (i.e., Type II error) than about the errors of falsely rejecting them (i.e., Type I error). In other words, the Agency is primarily concerned that the supporting test data provide sufficient assurance that each of the data quality requirements are met or exceeded. This assurance is related to the "power" (power is defined as 1 minus the Type II error) of the resulting test data to detect with known probability the likelihood that data quality would suffer with the proposed method. For instance, EPA may desire an 80 percent probability of rejecting H_3 if the proposed (i.e., new) test method has a 50 percent (or worse) loss in precision. In this example, the 50 percent precision loss is the change in data quality to be concerned about and the power (likelihood of correctly identifying it) is 80 percent.

The petitioner should also note that if the comparative approach is elected, then the approved test method provides the standard. The proposed method fails the equivalency test if the difference between its bias and the bias of the approved method is found to be significant. Where it is believed that the proposed method is more accurate than the approved method, or where resources necessary to apply the approved method are limited, then the absolute approach may be selected.

As noted earlier, both the significance level and the power of hypothesis tests about bias and precison are themselves data quality objectives. These measures describe the quality of the data submitted in support of the petition. Studies should have adequate power to detect negative changes in data quality that are of concern to EPA. Test power requirements for various data quality parameters should be determined on a case-by-case basis through discussions with the OSW program staff.

It may not be necessary to design a special experiment to support the petition. The petitioner may have extensive experience with the proposed test method for relevant waste samples. For example, data from a properly conducted quality control (QC) program may represent an acceptable alternative to a special experiment. Specifically, spiked waste sample data spanning a sufficiently long time period may provide the information needed to assess method precision and bias. The need for comparability of the petitioner's samples with those analyzed under SW-846 is a critical issue in determining whether the petition may be based solely on QC data.

Both the absolute and comparative cases can be further classified according to the need for multiple waste sites in the equivalency test program. The extent to which the proposed test method is used (e.g., local versus regional application) is a determining factor for requiring multiple sites. The petitioner should determine, in coordination with the OSW staff, which of the following categories best describes the nature of the equivalency study method:

a. Absolute - single site/local significance

b. Comparative - single site/local significance

c. Absolute - multiple sites/regional or national significance

d. Comparative - multiple sites/regional or national significance.

The first step in choosing an experimental design is to define the target population to which statistical inferences are desired. The design should incorporate relevant sources of variation and exclude irrelevant sources. One consequence of this basic principle is that if validation or equivalency data involve only one laboratory (or only one site), then approval may be limited to use of the method at that laboratory (or at that site). Equivalency studies for multilaboratory or multisite applications will usually require data from multiple laboratories or multiple sites.

Appendix B, Section 1 provides guidance for developing and conducting test programs and associated statistical analyses for petitioners who only plan to use categories (a) and (b) above.

Appendix B, Section 2 discusses issues related to petitions where multiple sites and multiple laboratories will use a test method. The development and analysis of test programs for these cases are significantly more complex than for the single site case; for this reason, in multisite or multilaboratory categories, it is strongly recommended that the services of a professional statistician be available to ensure an adequate and cost-effective test program.

Specific experimental designs are proposed (in Appendix B, Sections 1 and 2) for each of the four cases. These designs have been evaluated by EPA, and are considered minimal. Deviations from these designs would normally be in the direction of more data, e.g., more sites, more samples/days of analysis, and/or more concentrations. While replication is expected in equivalency testing, the petitioner is not encouraged to run more than two replicates.

SECTION 5

HOW TO PREPARE A PETITION

5.1 PURPOSE OF A PETITION

The purpose of a test method equivalency petition is to demonstrate to EPA that a proposed test method (whether it be a sampling method, waste characterization method, or any other method believed comparable to those found in SW-846) is equal or superior to the corresponding SW-846 method. To accomplish this, the petitioner must provide EPA with sufficient information to determine equivalency by following the approach outlined in 40 CFR 260.20 and 260.21. To identify what information must be provided, the petitioner should ask himself the following questions:

- What is the objective of the petition, i.e., can I provide a clear statement of scope and applicability?
- What information must be known about the proposed test method to understand how to use it?
- What test method performance information can be provided to demonstrate that it is equal to or better than an SW-846 method?

5.2 CONTENT OF A PETITION

Although the regulations do not require a specific format for the petition, for ease of review, the necessary information should be organized in a logical manner. Figure 5-1 is a suggested format for a petition. By following the suggested format and addressing the items in sufficient detail, the petitioner can contribute to a more expeditious and uniform evaluation of the test method petition.

5.2.1 Name and Address of Petitioner

The petition should begin with the following administrative information:

- Name of the firm submitting test method petition.
- Address (i.e., street, city, State, and zip code).
- Names, titles, and telephone numbers of persons to contact for additional petition information.

TEST METHOD EQUIVALENCY PETITION

1. Name and Address of the Petitioner

2. Certification of Accuracy and Responsibility

3. Description of Proposed Action

   Description of Test Method

   Description of Applicable Samples or Matrices

   Assessment of Limiting or Interfering Factors

   Test Method Quality Control Procedures

4. Statement of Need and Justification for the Proposed Test Method

   Need

   Justification

Figure 5-1. Suggested petition format.

5.2.2 Certification of Accuracy and Responsibility

Each petition must include the following certification statement:

I certify under penalty of law that I have personally examined and am familiar with the information submitted in this demonstration and all attached documents and that based on my inquiry of those individuals immediately responsible for obtaining the information, I believe that the submitted information is true, accurate, and complete. I am aware that there are significant penalties for submitting false information, including the possibility of fine and imprisonment.

Signed,

_____

Title _____

Date _____

5.2.3 Description of Proposed Action

This section of a petition contains descriptive information regarding (1) the test method itself, (2) the applicability of the test method, (3) any factors that limit or interfere with the performance of the method, and (4) the necessary quality control procedures that should accompany the method's use. In general, this section should contain all of the information necessary for EPA to understand the proposed method.

5.2.3.1 Test Method Description

The description of the proposed test method should be presented in a fashion that will help EPA evaluate the petition effectively and allow direct comparison with the current test method description. Therefore, descriptions of a proposed test method should follow the established format for the comparable SW-846 test method. The following sections reflect the current format and information needs for analytical test methods and RCRA waste characteristic test methods found in SW-846. In general, the test method description should contain all of the information described in these sections that is needed for any qualified person to perform the test method.

Scope and Application. This portion of the test method description identifies what the method accomplishes; i.e., it identifies the sample property or constituent(s) that the method measures. It also briefly describes the types of samples that one can analyze with acceptable accuracy. The petitioner should mention (1) steps taken to prepare the sample before the method is applied and (2) procedural changes required for special waste samples. Finally, the petitioner should identify any special qualifications required for personnel that will perform the test method.

Summary. Petitioners should provide a one-or two-paragraph summary of the test method description. If detection limit(s) are applicable to the test method, they should be provided for the properties or constituents that the method measures.

Interferences. Caused by unexpected chemical reactions or physical properties such as heat, light, or the physical state of a sample, interferences can lead one to believe that sample concentrations are higher or lower than actual. Therefore, it is important to identify and correct all interferences that may occur when the proposed test method is conducted. For example, in atomic absorption spectroscopy, light scattering and sample viscosity may cause interferences. With gas chromatography, sample processing hardware, glassware, low-purity solvents, or reagents could cause artifacts and/or elevated baselines in chromatographs. For ignitability test methods, ambient pressure or drafts could significantly affect flash point values.

Materials and Apparatus. The petitioner must provide a detailed list of materials required to conduct the test method procedures accurately. The apparatus must be described thoroughly, with product names and model numbers, equipment size, construction materials, and other information important for the success of the method identified. When an apparatus must be specially built for the test, the petitioner must include detailed construction instructions.

Reagents. A "reagent" is "a substance, chemical, or solution used in the laboratory to detect, measure, or otherwise examine other substances, chemicals or solutions"(5). All reagents used in any aspect of the test method (including sample preservation and preparation) must be identified in the description. Important details about reagents include:

- Proper grade(s) to use such as American Chemical Society (ACS)
- Concentrations of reagents and how to prepare them
- Maximum shelf life allowed for accurate analysis.

Sample Collection, Preservation, and Handling. Petitioners for test method equivalency must acknowledge that they use a sampling plan addressing considerations in Section 1 of SW-846. (This section concerns statistical techniques for obtaining accurate and precise samples.) This ensures that the sampling program does not interfere with the test method's performance.

Procedures. The test procedures are the heart of every test method and must be described in a succinct, stepwise fashion. Excluding petitions for new sampling methods, descriptions must include all procedures after sample collection. If the method requires sample preparation steps, the petitioner must include these in this section. The following topics are considered test method procedures:

- Sample preparation, e.g., extracting a constituent from a solid sample into a liquid media.

- Test method apparatus.
- Entering the sample into the testing device.
- Operating conditions for test method apparatus.
- Calibration procedures
  - number of standard calibration concentration levels to use per parameter
  - preparation method for blanks and calibration samples
  - frequency for preparing fresh blanks and calibration samples
  - tabulation of results from calibration standards
  - frequency of verifying instrument calibration
  - acceptable qualitative minimum recovery yields
  - amount of variation allowed from the calibration standard before recalibration is required.
- Analytical test procedures (where applicable)
  - retention times
  - sensitivities
  - acceptable recovery yields
  - method for handling unexpected events such as peak areas that are greater than the linear range of the analytical system
  - method for verifying the absence of interference
  - method for coping with interference.
- Calculating results
  - units of measure
  - complete equations
  - data validation, if applicable.
- Proper management of remaining sample and contaminated laboratory equipment and materials.

Quality Control (QC) Procedures. The information required in this section is described in "Test Method Quality Control Procedures" (Section 5.2.3.4).

References. Each test method description shall include a list of literature or other references (1) pertinent to the development and content of the method, and (2) demonstrating the procedure's applicability. A complete reference list allows the person performing the test method to refer to original documents if questions arise.

5.2.3.2 Description of Applicable Samples or Matrices

A test method equivalency petition should describe each sample type or matrix for which the proposed method is suitable. Information about the

sample's chemical and physical character provides both the person performing the test method and petition reviewer insight into the test method's applicability.

Where the method is applicable to a general type of matrix such as oily sludge, information about the percent water, solids contents, or any other physical property that affects the test method's performance facilitates the petition evaluation.

If the test method is designed for a specific waste stream(s), information about the variability of the waste stream as it might offset the applicability of the method is required.

In general, the petitioner should provide enough information about the sample(s) or matrix for the petition reviewer to evaluate the test method's applicability.

5.2.3.3 Assessment of Limiting or Interfering Factors

This section of the petition includes more than the "interference" statement in the method description. It is an assessment of how interferences and limitations affect the performance of the proposed method.

The petitioner must (1) identify every limitation and interference known to occur with the test method and (2) describe the effect each limitation or interference has on the proposed test method's results and present test data that demonstrate such impacts. Where these limitations and interferences depend on specific samples or matrices, a description for each interference or limitation should be provided. Where measures of waste constituents or waste characteristics typically vary, the description must include demonstrated or predicted effects which the limitations or interferences have on the accuracy of such variations. The petitioner should describe any corrective measures the analyst can take to prevent or minimize the limitation or interference.

A common test method limitation is the "quantitation limit" which represents the minimum waste constituent concentration or value of a waste property that a test method can measure accurately. When preparing this portion of the petition, the petitioner should identify the proposed method's known quantitation limit (where applicable) and discuss its implications for the types of samples intended for testing (i.e., how often the method will be able to detect the constituent or property of interest given a sample's typical characteristics and composition). Other limitations encountered with test methods include the type of matrices the method apparatus can accept and the species of a constituent that are detectable, such as the variety of halogenated compounds.

"Light scattering" when particulate matter is present in the sample is an example of test method interference. In atomic absorption spectroscopy, it alters the sample's true absorption of light for a specific metal. Therefore, a petitioner should describe how significantly this interference changes true absorption. Other examples of interference are high sample viscosity, test method apparatus construction materials, and low-purity reagents.

5.2.3.4 Test Method QC Procedures

Quality control procedures "define the frequency and methods of checks, audits, and reviews necessary to identify problems and dictate corrective action, thus verifying product quality" (SW-846). Quality control ensures that test method results are representative of the sample's true character. Procedures such as good recordkeeping, equipment maintenance, personnel training, and test method practices contribute to a successful QC program.

Each petition must describe clearly the QC procedures followed for the proposed test method because they will become part of the test method description if approved for SW-846. The petitioner also should describe how the selected QC procedures enhance the method's overall performance. QC procedures that should be addressed in the petition include:

- Training and evaluating personnel that operate test method equipment.
- Maintaining, inspecting, and servicing test method equipment.
- Test method procedures (where applicable)
  - preparing calibration curves for a blank and a specific number of standard solutions
  - diluting samples if the constituent concentration exceeds or falls on the plateau of the calibration curve for the highest concentration standard solution
  - running a given number of blanks for each sample batch to assess contamination
  - checking standard solutions and a duplicate sample after a certain increment of samples is tested
  - incorporating spiked or standard reference samples into the regimen periodically to ensure that test procedures are followed and test method equipment is operated correctly.
- Recordkeeping procedures
  - labeling laboratory samples and conducting a sample chain-of-custody program within the laboratory
  - using statistical procedures to check accuracy, precision and bias
  - continuously reviewing analytical results to identify problems
  - documenting training and equipment performance
  - maintaining all records properly.

The statistical check for accuracy, precision, and bias listed above requires additional discussion to understand fully the information requirements for a petition.

"Accuracy" measures how well a specific sample data point agrees with the true value of interest, and it is assessed based on test result precision and bias. "Precision" represents how well repeated measurements of the same constituent or property agree with one another, and it is assessed using the standard deviation of a series of controlled measurements. "Bias" is the constant difference between the average of data points produced by the test method and the true value of interest; thus, bias measures any systematic error in a test method and is calculated as the difference between the true constituent value and the average of laboratory runs. A petition should address these three statistical concepts along with steps to determine if a database is complete, representative, and comparable to other databases for the same test method.

Section 10 of SW-846, "Quality Control/Quality Assurance," illustrates how EPA defines and uses QC procedures to ensure that test methods are performed properly. Section 10 includes guidance on:

- Using QA/QC procedures to ensure achievement of program goals.
- Developing a sampling program that can be measured for how well samples represent the true value of interest.
- Developing a test program that provides data at the level of accuracy and precision that will be required by users of the data for decisionmaking under RCRA.
- Assessing the quality of the data that result from use of the test method, e.g., accuracy and precision.

Appendix C contains the full text of SW-846, Second edition, Section 10. The petitioner should, however, refer to the most current edition of SW-846 when preparing a petition.

5.2.4 Statement of Need and Justification for the Proposed Test Method

This final section of a petition contains two parts. In the first part, the petitioner should briefly state why the proposed method is needed. The second part (the justification) should contain the test data that will be used to establish equivalency between the proposed method and the existing SW-846 method. Thus, while the preceding section describes the test method and associated procedures and provides the information necessary to understand the method, this section supplies the data that will enable EPA to evaluate the method's performance equivalency.

5.2.4.1 Need

The petitioner shall describe briefly the benefits of the proposed test method compared to the existing method. For example, (1) the proposed method may be less expensive to run than the approved methods, or (2) the proposed method may employ a new, proprietary technique. Numerous other reasons could exist.

5.2.4.2 Justification--The Presentation of Test Data

Petitioners who request approval of a proposed test method are required to provide performance data and statistical analyses that are used to demonstrate the equivalency of the proposed method to the approved SW-846 procedure. Along with the test data, the petitioner should provide all QC data to substantiate the validity of the test data. Petition approval is then based on EPA's evaluation of the petitioner's data and statistical analyses from representative samples of hazardous wastes, ground water, or whatever the appropriate matrix may be.

Where the proposed test method measures waste characteristics or constituents, EPA requires that petitioners propose an experimental design for the Agency's review and approval before testing begins. (See Figure 3-1.) This design must yield the comparability data that reflect the results of testing samples spanning the range of the method's applicability. This design includes both the sampling and testing program to be conducted for the petition. The number of samples and their locations may vary with each petition along with the number of sample replicates and the timing for conducting tests. Section 4 and Appendix B discuss the development of simple and complex experimental designs.

Sampling Procedures. Except when new sampling techniques are proposed, petitioners may obtain the samples in any manner so long as the sample is representative of the material that is to be tested. If the ambient level of the constituent or property of interest is below the detection limit of the SW-846 approved test procedure, the samples should be spiked to levels near the detection limits of the SW-846 approved test procedure and the recommended maximum property or concentration level.

Documentation of Test Results. All test results should be documented with the following information:

- Name, address, and telephone number of each laboratory facility performing the testing, if different from the sampling laboratory
- Date the test method was performed
- Sample number
- Parameter or constituent measured
- Test method
- Test results.

Explanations or additional information on inconsistencies or deviations in test results should be furnished as necessary. A suggested format for displaying the test result information is shown as Table 3-1A in the Example Test Method Equivalency Petition (Appendix D).

Statistical Testing. The petitioner performs the statistical analyses of all test results and submits these analyses along with test data to EPA as part of the petition. The data analyses entail a series of statistical tests that are used to determine if results from the proposed method satisfy the data quality objectives established at the beginning of the petitioning process. Section 4 and Appendix B of this manual provide more detailed information about statistical evaluation procedures.

SECTION 6

CHECKLIST OF PETITION REQUIREMENTS

Table 6-1 is a checklist of items developed to help petitioners prepare a complete petition; the checklist is designed so the petitioner can determine whether an item is or is not properly addressed. By addressing the checklist items appropriately, the petitioner can reduce the likelihood that EPA will request additional information about the method.

The checklist is divided into the four major categories discussed in Section 5, "How to Prepare a Petition":

- Name and Address
- Certification of Accuracy and Responsibility
- Description of Proposed Action
- Statement of Need and Justification for the Proposed Test Method.

These four categories also correspond to the example equivalency petition presented in Appendix D of this manual. Applicable RCRA regulations are cited within the checklist.

TABLE 6-1. A CHECKLIST OF ITEMS TO INCLUDE IN A PETITION

I. NAME AND ADDRESS (40 CFR 260.20 (b)(1))

a. Are the petitioning party's name and address identified? __yes __no

b. Is a specific contact listed along with his or her title and telephone number? __yes __no

II. CERTIFICATION OF ACCURACY AND RESPONSIBILITY

a. Does the petition include the appropriate certification statement? __yes __no

III. DESCRIPTION OF PROPOSED ACTION (40 CFR 260.20 (b)(3))

a. Does the petition include a test method description that follows SW-846 format? (40 CFR 261.21 (b)(1))? __yes __no

i. Does the "Scope and Application" section --

- address sample properties or constituents measured
- describe types of samples to be tested
- briefly describe sample preparation before the test method is applied
- specify any procedural changes for special samples
- specify qualifications for participating personnel? __yes __no

ii. Does the "Summary of Method" section --

- summarize the test method description
- include any detection limits? __yes __no

iii. Does the "Interference" section identify all potential interferences that may occur when the test method is conducted? For example -- __yes __no

- light scattering
- nonspecific absorption
- high concentrations of other metals suppressing absorbance
- viscosity variations that may alter aspiration rates
- artifacts from sample processing glassware, hardware, or low-purity reagents
- additional sample cleanup for desired sensitivity
- contaminated glassware
- plastic apparatus
- ambient pressure
- sample nonhomogeneity
- operator bias

(Continued)

TABLE 6-1. A CHECKLIST OF ITEMS TO INCLUDE IN A PETITION (Continued)

III.a.iv. Does the "Apparatus and Materials" section --

- provide a detailed list of each item used for the method
- thoroughly describe apparatus as needed
- provide detailed construction guidelines for apparatus that must be built?

__yes __no

v. Does the "Reagents" section identify reagents according to --

- test method
- sample preservation and preparation
- proper grades
- reagent concentrations and preparation method
- maximum reagent shelf life allowed?

__yes __no

vi. Does the "Sample Collection, Preservation, and Handling" section

- address the considerations in Section 1 of SW-846

__yes __no

vii. Does the test method "Procedures" section --

- describe procedures in a succinct, stepwise fashion
- begin immediately after sample collection
- address the following topics as appropriate:
  - sample preparation
  - entering the sample into the testing device
  - operating conditions of the test method apparatus
  - calibration procedures
  - test method procedures
  - calculation of results
  - management of leftover samples and contaminated equipment and materials?

__yes __no

viii. QC procedures (see Item III.d)

ix. Does the test method description include a list of references --

- pertinent to the development and content of the method
- demonstrating the procedure's applicability?

__yes __no

(Continued)

TABLE 6-1. A CHECKLIST OF ITEMS TO INCLUDE IN A PETITION (Continued)

III.b. Does the petition include a section on "Description of Applicable Samples or Matrices" (40 CFR 261.21 (b)(2)) that --

- for a general waste matrix, such as sludge, describes the physical properties affecting test method performance
- for specific waste streams, provides RCRA waste numbers and a brief background on the waste generation source
- in general, provides enough information about the waste for EPA to evaluate the method's applicability? __yes __no

c. Does the petition include a section on "Assessment of Limiting or Interfering Factors" (40 CFR 260.21 (b)(4)) that addresses--

- identification of interferences and limitations
- effect of interferences and limitations on test method performance (effects on specific wastes where appropriate)
- corrective measures available __yes __no

d. Does the petition include a section on "Test Method Quality Control Procedures" (40 CFR 260.21(b)(5)) that addresses --

- training and evaluation of laboratory personnel
- maintenance of equipment
- test method practices
- recordkeeping and statistical procedures __yes __no

IV. STATEMENT OF NEED AND JUSTIFICATION FOR THE PROPOSED TEST METHOD (40 CFR 260.20 (b)(4))

a. Does the petitioner briefly describe why the proposed test method should be approved? __yes __no

b. Does the petition include an experimental design that satisfies the data quality objectives established by the EPA? __yes __no

c. Does the petition include the test data (absolute or comparative) required as a result of prepetition negotiations with EPA? __yes __no

d. Are test data displayed clearly with full documentation for each sample? __yes __no

e. Does the petition present complete statistical analyses of test data along with QC data collected during testing? __yes __no

REFERENCES

1. Test Methods for Evaluating Solid Waste: Physical/Chemical Methods. SW-846, 3rd Edition, U.S. Environmental Protection Agency, Washington, D.C. 1986. Available from Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402 as document 955-001-00000.

2. Guidelines for Selection and Validation of US EPA's Measurement Methods. Draft document prepared by Office of Acid Deposition, Environmental Monitoring and Quality Assurance, U.S. Environmental Protection Agency, Washington, DC 20460. January 1986.

3. Validation of Testing/Measurement Methods. Prepared by U.S. Environmental Protection Agency, Office of Research and Development for the Office of Solid Waste, Washington, DC 20460. EPA 600/X-83-060.

4. Harmonization of Biological Testing Methodology: A Performance-based Approach. Aquatic Toxicology and Hazard Assessment: Eighth Symposium. ASTM STP891. R.C. Bahner and D.J. Hamsen, editors. American Society for Testing and Materials. Philadelphia, PA. 1985. pp. 288-301.

5. MacGraw-Hill Dictionary of Scientific and Technical Terms, 2nd Edition. MacGraw-Hill Book Company, New York City, NY. 1978.

APPENDIX A

40 CFR 260, SUBPART C - RULEMAKING PETITIONS

APPENDIX A

40 CFR 260, SUBPART C - RULEMAKING PETITIONS

260.20 General

(a) Any person may petition the Administrator to modify or revoke any provision in Parts 260 through 265 of this chapter. This section sets forth general requirements which apply to all such petitions. Section 260.21 sets forth additional requirements for petitions to add a testing or analytical method to Part 261, 264, or 265. Section 260.22 sets forth additional requirements for petitions to exclude a waste at a particular facility from 261.3 of this chapter or the lists of hazardous wastes in Subpart D of Part 261[1].

(b) Each petition must be submitted to the Administrator by certified mail and must include:

(1) The petitioner's name and address;

(2) A statement of the petitioner's interest in the proposed action;

(3) A description of the proposed action, including (where appropriate) suggested regulatory language; and

(4) A statement of the need and justification for the proposed action, including any supporting tests, studies, or other information.

(c) The Administrator will make a tentative decision to grant or deny a petition and will publish notice of such tentative decision, either in the form of an advanced notice of proposed rulemaking, a proposed rule, or a tentative determination to deny the petition, in the Federal Register for written public comment.

(d) Upon the written request of any interested person, the Administrator may, at his discretion, hold an informal public hearing to consider oral comments on the tentative decision. A person requesting a hearing must state the issues to be raised and explain why written comments would not suffice to communicate the person's views. The Administrator may in any case decide on his own motion to hold an informal public hearing.

(e) After evaluating all public comments the Administrator will make a final decision by publishing in the Federal Register a regulatory amendment or a denial of the petition.

---

[1] Note: Section 260.22 (Delisting Procedures) is not applicable to this guidance manual.

260.21 Petitions for equivalent testing or analytical methods

(a) Any person seeking to add a testing or analytical method to Part 261, 264, or 265 of this chapter may petition for a regulatory amendment under this section and 260.20. To be successful, the person must demonstrate to the satisfaction of the Administrator that the proposed method is equal to or superior to the corresponding method prescribed in Part 261, 264, or 265 of this chapter, in terms of its sensitivity, accuracy, and precision (i.e., reproducibility).

(b) Each petition must include, in addition to the information required by 260.20 (b):

(1) A full description of the proposed method, including all procedural steps and equipment used in the method;

(2) A description of the types of wastes or waste matrices for which the proposed method may be used;

(3) Comparative results obtained from using the proposed method with those obtained from using the relevant or corresponding methods prescribed in Part 261, 264, or 265 of this chapter;

(4) An assessment of any factors which may interfere with, or limit the use of, the proposed method; and

(5) A description of the quality control procedures necessary to ensure the sensitivity, accuracy, and precision of the proposed method.

(c) After receiving a petition for an equivalent method, the Administrator may request any additional information on the proposed method which he may reasonably require to evaluate the method.

(d) If the Administrator amends the regulations to permit use of a new testing method, the method will be incorporated in "Test Methods for the Evaluation of Solid Waste: Physical/Chemical Methods," SW-846, U.S. Environmental Protection Agency, Office of Solid Waste, Washington, D.C. 20460.

APPENDIX B

STATISTICAL PROCEDURES TO EVALUATE TEST METHOD EQUIVALENCY

APPENDIX B

STATISTICAL PROCEDURES TO EVALUATE TEST METHOD EQUIVALENCY

This appendix is intended to assist both EPA and the petitioner in the selection and evaluation of experimental designs and methods of statistical analysis for equivalency studies. Section 1 establishes the basic approaches to equivalency testing and addresses simple (single-site) cases. Sections 2 and 3 deal with more complex (multi-site) cases. Although this appendix is written mainly for statisticians, much of the material is elementary; it is included here to fix ideas, terminology, and approaches recommended by EPA. Alternative designs and approaches are of course possible and may be negotiated between the Agency and the petitioner.

B1. EXPERIMENTAL DESIGNS AND METHODS OF DATA ANALYSIS FOR SINGLE SITE CASES

This section presents statistical methods to determine the equivalency of test methods proposed for simple (single-site) cases. To assist the readers, a list of symbols used in Section B1 is provided at the end of Appendix B.

B1.1 Experimental Designs for Single-Site Cases

Ideally, the choice of experimental design and statistical analysis for equivalency testing involves a professional statistician. As noted earlier, EPA encourages the petitioner to use the services of a statistician for those cases involving multiple waste site designs and analyses. For single site cases, the following sections offer some reasonably simple experimental designs and statistical analyses where professional statistical assistance may not be available.

The designs for the absolute and comparative cases are quite similar. Both involve a data matrix consisting of two columns (corresponding to levels of spiking concentrations for the absolute case and test methods for the comparative case) and ten rows (corresponding to the number of days). Each

cell of the data matrix contains two observations or replicate measurements. A schematic of these designs is given in Figure B1-1. Each design uses ten randomly selected waste samples from a single site. For the absolute case, each sample is split into four subsamples, two subsamples are spiked at the low concentration and two are spiked at the high concentration. Spiking waste samples yields information on precision and bias. The spiking concentrations should be near the regulated level of the waste component and also in the operation range of the method where the percent recovery is relatively constant. In particular, the spiking concentrations should be above the detection limit, with recommended spiking concentrations of T/2 and 3T/2, where T is the regulatory threshold.

For the comparative case, each sample is split into four subsamples, two subsamples (randomly selected) are analyzed by the proposed test method and the remaining two are analyzed by the approved method. All four subsamples are to be analyzed on a single day. These designs are regarded as minimal. If additional resources are available, increasing the number of concentration levels, the number of days and/or the number of samples is recommended.

B1.2 Preliminary Analysis

This section gives procedures for (1) identifying and handling unusual or "outlying" measurements, (2) testing whether error variances are equal (as assumed when applying the analysis of variance procedure), and (3) determining what data transformation will help stabilize the variance when unequal variances are found. Material in this section applies to both the absolute and comparative single-site cases.

B1.2.1 Screening for Outliers. Whenever the numerical value of an observation (as opposed to the rank, for instance) is used in an analysis, outliers or unusual observations can seriously affect the results. Therefore,

a) Absolute

	Spiked Concentration: Low	High
Day 1 (Sample 1)	x x	x x
Day 2 (Sample 2)	x x	x x
Day 3 (Sample 3)	x x	x x
Day 4 (Sample 4)	x x	x x
Day 5 (Sample 5)	x x	x x
Day 6 (Sample 6)	x x	x x
Day 7 (Sample 7)	x x	x x
Day 8 (Sample 8)	x x	x x
Day 9 (Sample 9)	x x	x x
Day 10 (Sample 10)	x x	x x

b) Comparative

	Test Method: Proposed	Approved
Day 1 (Sample 1)	x x	x x
Day 2 (Sample 2)	x x	x x
Day 3 (Sample 3)	x x	x x
Day 4 (Sample 4)	x x	x x
Day 5 (Sample 5)	x x	x x
Day 6 (Sample 6)	x x	x x
Day 7 (Sample 7)	x x	x x
Day 8 (Sample 8)	x x	x x
Day 9 (Sample 9)	x x	x x
Day 10 (Sample 10)	x x	x x

Figure B1-1. Layout of experimental designs for single site case.

prior to running the planned analysis, one should carefully examine the data and identify the "suspect" values for further study. If it is determined (based on supportable evidence) that an outlying observation is a gross deviation from prescribed procedures (e.g., recording or calculation errors), the value in question should be rejected unless it can be restored through a correction to the identified error. In those cases where there is not sufficient evidence to make this determination, a statistical procedure should be employed in identifying "outliers". Without some objective procedure, one may find experimenters using analytical results to determine whether unusual values are to be retained or discarded. Clearly, this should be avoided.

It is suggested that each column of the designs shown in Figure B1-1 be screened for outliers, i.e., each concentration in the absolute case or each method in the comparative case. The discussion here will be in terms of concentration for the sake of specificity, but can be adapted to the comparative case by substituting method for concentration. For each concentration, calculate the overall mean ($\overline{Y}$), the between-day mean square error (MSB), and the within-day mean square error (MSW). Procedures for estimating MSB and MSW are given in Section B1.3; they may also be found in most statistical texts under discussions of one-way analysis of variance. The total variance of a single observation from Figure B1-1 is estimated by

$$s^2_{TOT} = \tfrac{1}{2}(MSW + MSB)$$

and the corresponding standard deviation is

$$\sqrt{s^2_{TOT}} = \sqrt{\tfrac{1}{2}(MSW + MSB)}$$

It is also suggested that any observation more than 4.0 standard deviations from the grand average, i.e., any value outside $\overline{Y} \pm 4.0\ s_{TOT}$, be considered

suspect. (An alternative is to use the standard deviation derived from the twenty observations for a given concentration, ignoring days, as an estimate of s_{TOT}.)

The petitioner is free to suggest other procedures that may be used to identify outliers such as Thompson's t-test or the Dixon ratio test. These and other approaches for handling outlying observations are described in (1).

In order to preserve the balance of an experimental design, observations rejected as outliers should be replaced by representative values. A simple approach is to substitute the overall mean, excluding the suspect observation for the value of the outlier. Other procedures may be employed. In all cases, the raw data should be reported along with an indication of what observations were considered outliers, how they were detected, and how they were treated.

B1.2.2 Equality of Replicate Variance. One of the assumptions for applying the analysis of variance procedure is that the error variances (or replication variances) are equal under all conditions. The experimental designs shown in Figure B1-1 include duplicate measurements for each condition, i.e., two observations within each cell. There are two reasons for utilizing replicate measurements. First, it enables the assumption of equal variances to be tested, and, second, it allows the concentration by day interaction effect and the method by day interaction effect to be evaluated.

It is suggested that the initial statistical analyses use recovery of the spiked amount (i.e., difference in measured concentrations before and after spiking divided by spiked concentration) as the response variable in the absolute case and the logarithm of the observed measurement as the response variable in the comparative case. If waste samples contain concentrations in the working range of the test methods, one often finds the standard deviation

of concentration measurements to be proportional to concentration. When this is true, a benefit of analyzing recoveries in the absolute case and logarithms of measurements in the comparative case is that variances will be approximately equal.

To test for equality of replicate variances, calculate (for the i^{th} day and j^{th} concentration or method) the within-cell average ($\overline{Y}_{ij}$) and standard deviation (S_{ij}) (of recoveries in the absolute case; of logarithms of measurements in the comparative case) for each cell. Plot $\log(S_{ij})$ vs. $\log(\overline{Y}_{ij})$ and use ordinary linear regression

$$\log (S_{ij}) = a \log(\overline{Y}_{ij}) + b$$

to test for dependence of variation on level of concentration. If the regression is significant, i.e., coefficient "a" is significantly different from zero, the variances are considered unequal and some data transformation should be employed in an effort to stabilize the variances before applying the analysis of variance procedure. One approach is to use the power transformation, Y^{1-a}, on the raw data and test again for equality of variances. Note that "a" is the coefficient in the above regression equation. When a=1, the log transformation is appropriate (2).

B1.3 Statistical Analysis for Single-Site Cases With Absolute Data Quality Objectives

This section provides details of the statistical test used to evaluate whether a proposed test method is at least equivalent (in terms of precision and bias) to the absolute data quality objectives specified by the EPA for single-site cases. It is assumed that preliminary analyses, i.e., tests for outliers and variance described in Section B1.2, have already been carried out. These objectives are formulated in terms of the null hypotheses (H_1 and H_2) stated in Section 4.2. All analyses discussed here are to be performed

separately for each level of concentration. If any test leads to rejection of either H_1 or H_2, then the proposed method is considered unacceptable. In addition, it is assumed that two replicate measurements are taken for each sample and that any modificaton on the minimal experimental design is in the direction of additional concentrations and/or days. The formulas provided in the analysis of variance (ANOVA) table have been simplified for the case of two replicates per sample. All the analyses described below are dependent on values of either the F, t, or chi-square distributions. In situations where this design has been modified to include additional days/samples, appropriate percentiles can be found in Table B1-1 for up to 20 days (the number of data nts, N, equals 40).

Once the data have been collected, it should be presented in a manner similar to the configuration displayed in Table B1-2. Here the first column shows the day (i) that the sample was tested, and the second and third columns contain recoveries (Y_{1i} and Y_{2i}) obtained from the two replicate measurements. The remaining columns are simple manipulations of the recoveries which are used in the calculations for the ANOVA table shown in Table B1-2(b). Note that totals are calculated for some of the columns in Table B1-2(a); these totals are referred to by a capital letter for easy reference in later formulas. The computations for the ANOVA table should be straightforward if the data are arranged and manipulated as described here. However, it is suggested that the petitioner work through the example provided at the end of this section before beginning to analyze his own data.

After the ANOVA table has been completed, the petitioner should screen for outliers as described in Section B1.2.1. Note that the between-day and

TABLE B1-1. PERCENTILES OF THE t, χ^2 AND F DISTRIBUTIONS

Percentiles of the t and χ^2 distributions			Percentiles of the F-distribution				
Degrees of Freedom	$t_{.975}$	$\chi^2_{.95}$	Numerator Degrees of Freedom	Denominator Degrees of Freedom	$F_{.90}$	$F_{.95}$	$F_{.975}$
9	2.26	16.92	1	29	2.89	4.18	5.59
10	2.23	18.31		30	2.88	4.17	5.57
				40	2.84	4.08	5.42
11	2.20	19.68	9	9	2.44	3.18	4.03
12	2.18	21.03		10	2.35	3.02	3.78
13	2.16	22.36		20	1.96	2.39	2.84
14	2.14	23.69	10	10	2.32	2.98	3.72
15	2.13	25.00		11	2.25	2.85	3.53
				22	1.90	2.30	2.70
16	2.12	26.30	11	11	2.23	2.82	3.47
17	2.11	27.59		12	2.17	2.72	3.32
18	2.10	28.87		24	1.85	2.22	2.59
19	2.09	30.14	12	12	2.15	2.69	3.28
20	2.09	31.41		13	2.10	2.60	3.15
				26	1.81	2.15	2.49
21	2.08	32.67	13	13	2.08	2.58	3.12
22	2.07	33.92		14	2.04	2.51	3.01
23	2.07	35.17		28	1.77	2.09	2.41
24	2.06	36.42	14	14	2.02	2.48	2.98
25	2.06	37.65		15	1.99	2.42	2.89
				30	1.74	2.04	2.34
26	2.06	38.89	15	15	1.97	2.40	2.86
27	2.05	40.11		16	1.94	2.35	2.79
28	2.05	41.34		32	1.71	1.99	2.28
29	2.05	42.54	16	16	1.93	2.33	2.76
30	2.04	43.77		17	1.90	2.29	2.70
				34	1.68	1.95	2.22
35	2.03	49.80	17	17	1.89	2.27	2.67
40	2.02	55.76		18	1.86	2.23	2.62
				36	1.66	1.92	2.17
			18	18	1.85	2.22	2.60
				19	1.83	2.18	2.55
				38	1.63	1.88	2.13
			19	19	1.82	2.17	2.53
				20	1.80	2.14	2.48
				40	1.61	1.85	2.09

USWER Policy Directive #9433.00-2

Table B1-2. LAYOUT FOR SINGLE SITE CASE WITH ABSOLUTE DQOs

a. Data Configuration for One Concentration

Day	Recoveries					
i	Y_{1i}	Y_{2i}	Y_{1i}^2	Y_{2i}^2	$T_i = Y_{1i} + Y_{2i}$	T_i^2
1	Y_{11}	Y_{21}	Y_{11}^2	Y_{21}^2	T_1	T_1^2
2	Y_{12}	Y_{22}	Y_{12}^2	Y_{22}^2	T_2	T_2^2
3	Y_{13}	Y_{23}	Y_{13}^2	Y_{23}^2	T_3	T_3^2
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
D	Y_{1D}	Y_{2D}	Y_{1D}^2	Y_{2D}^2	T_D	T_D^2
Totals			$G = \sum_{i=1}^{D} Y_{1i}^2$	$H = \sum_{i=1}^{D} Y_{2i}^2$	$J = \sum_{i=1}^{D} T_i$	$K = \sum_{i=1}^{D} T_i^2$

b. Analysis of Variance (ANOVA) Table for One Concentration

Source	Degrees of Freedom	Sum of Squares	Mean Square Error	F-statistic
Between Days	D-1	$SSB = (K/2)-(J^2/N)$	$MSB = SSB/(D-1)$	$F = MSB/MSW$
Within Days	N-D	$SSW = SST-SSB$	$MSW = SSW/(N-D)$	
Total	N-1	$SST = G + H - (J^2/N)$		

$N = 2D$ (total number of observations)

$\overline{Y} = J/N$ (overall mean recovery)

$s^2 = SST/(N-1)$ (sample variance)

within-day mean square errors (MSB and MSW) have already been calculated for the ANOVA table, and the formula for the overall mean ($\overline{Y}$) is given at the bottom of Table B1-2. As mentioned before, any outlier should either be corrected (if an error was discovered) or replaced by a representative value in order to preserve the balance of the design. In either case, all calculations performed up to this point will have to be repeated with the new value substituted.

Before the system precison and bias can be tested, it is necessary to determine if there is a significant day effect (i.e., the average recovery varies systematically over the days). This is done using the F-statistic calculated in the ANOVA table. If this value is greater than or equal to the critical F-value, then it is concluded that there is a significant day effect. The critical F-value for this test is $F_{D-1,N-D,.90}$ where (D-1) and (N-D) are the numerator and denominator degrees of freedom, respectively, and .90 is the level of confidence for the test. For the minimal design, there are 10 days (D) with 2 replicates per day, yielding a total sample size N = 2D = 20. The critical F-value is $F_{9,10,.90} = 2.35$. Therefore, under the minimal design, if the F-value from the ANOVA table is greater than or equal to 2.35, it is concluded that a significant day or sample effect is present. The result of this test determines the approach to be used in the following analyses of system bias and precision.

Case 1--Day Effect Not Significant

If the day effect is not significant (i.e., F-statistic < 2.35), bias and precision can be tested using the overall mean $\overline{Y}$ and the sample variance s^2 (from the bottom of Table B1-2). A 95 percent confidence interval (CI) for the true mean recovery, μ, can be computed using the following formula:

$$Pr(\overline{Y} - t_{N-1,.975}\sqrt{\frac{s^2}{N}} \leq \mu \leq \overline{Y} + t_{N-1,.975}\sqrt{\frac{s^2}{N}} = .95$$

where $t_{N-1,.975}$ is the 97.5th percentile of the t-distribution with N-1 degrees of freedom. When N=20 (minimal design), $t_{19,.975} = 2.09$ and the above CI simplifies to

$$Pr(\overline{Y} - 2.09\sqrt{\frac{s^2}{20}} \leq \mu \leq \overline{Y} + 2.09\sqrt{\frac{s^2}{20}} = .95$$

$$Pr(\overline{Y} - .467s \leq \mu \leq \overline{Y} + .467s) = .95$$

Substituting values for $\overline{Y}$ and s will yield lower and upper bounds of μ for the proposed method. In the absolute case, the EPA will specify a maximum bias (b_o) for the proposed method. Since, in the absence of bias, recovery is defined to be 1, the Agency will accept a proposed method whose true mean recovery is somewhere between ($1-b_o$, $1+b_o$). Therefore, if any portion of the 95 percent CI for μ overlaps the interval ($1-b_o$, $1+b_o$), then the bias of the proposed method is considered to be within acceptable limits. Similarly, a lower bound can be given for the true variance, σ^2, of the proposed method (still assuming no day effect)

$$Pr\left[\frac{(N-1)s^2}{\chi^2_{N-1,.95}} \leq \sigma^2\right] = .95$$

where $\chi^2_{N-1,.95}$ is the 95th percentile of the chi-square distribution with N-1 degrees of freedom. For the minimal design $\chi^2_{19,.95} = 30.14$. The Agency will also specify a maximum variance, σ^2_o, for a proposed method. If the lower bound of the true variance is less than or equal to σ^2_o, then the precision of the proposed method is considered to be within acceptable limits.

Case 2--Day Effect Significant

If the day effect is significant (i.e., F-statistic $\geq$ 2.35), methods for testing bias and precision are slightly different. In calculating the CI for the true mean recovery, μ, s^2 should be replaced by MSB (from the ANOVA table) and the degrees of freedom of the t-statistic is changed from (N-1) to (D-1) yielding:

$$Pr(\bar{Y}-t_{D-1,.975}\sqrt{\frac{MSB}{N}} \leq \mu \leq \bar{Y}+t_{D-1,.975}\sqrt{\frac{MSB}{N}} = .95$$

For the minimal design, $t_{9,.975} = 2.26$. As noted for the case where there was no day effect, if this interval overlaps the interval $(1-b_o, 1+b_o)$, then the hypothesis H_1: $|b_{new}| \leq b_o$ is not rejected, and the proposed method is considered acceptable. In order to test the precison when there is a day effect, the following calculations must be made:

$$g = s^2_{TOT} = 1/2(MSB + MSW) ,$$

$$n = \frac{4g^2}{\frac{MSB^2}{D-1} + \frac{MSW^2}{N-D}} .$$

The value n is an approximation for the new degrees of freedom and should be rounded up to the next highest integer (e.g., if n = 11.21, then use n = 12). A 95 percent CI for the lower bound of the true variance of the proposed method is

$$Pr\left[\frac{ng}{\chi^2_{n,.95}} \leq \sigma^2\right] = .95$$

As before, if the lower bound is less than or equal to the σ^2_o specified by the EPA, then the hypothesis H_2: $\sigma^2_{new} \leq \sigma^2_o$ is not rejected, and the precision of

the proposed method is considered acceptable. Note that the degrees of freedom for the chi-square statistic is n rather than (N-1), as in the case of no day effect. This percentile will have to be looked up in Table B1-1 even for the minimal design, since the degrees of freedom is a function of the data.

Numerical Example 1

Hypothetical recovery data from waste samples spiked at two concentration levels are analyzed to illustrate the calculations involved in applying some of the analytical procedures just described.

Example 1. Analysis of Recoveries From Low and High Concentrations For The Single Site Case With Absolute DQOs

A. Low Concentration

Day	Recoveries					
i	Y_{1i}	Y_{2i}	Y_{1i}^2	Y_{2i}^2	$T_i = Y_{1i} + Y_{2i}$	T_i^2
1	1.17	1.24	1.37	1.54	2.41	5.81
2	1.17	1.66	1.37	2.76	2.83	8.01
3	0.86	0.40	0.74	0.16	1.26	1.59
4	1.00	0.39	1.00	0.15	1.39	1.93
5	0.70	0.48	0.49	0.23	1.18	1.39
6	1.43	0.41	2.04	0.17	1.84	3.39
7	0.98	1.60	0.96	2.56	2.58	6.66
8	1.66	0.74	2.76	0.55	2.40	5.76
9	0.96	1.08	0.92	1.17	2.04	4.16
10	1.17	1.36	1.37	1.85	2.53	6.40
Totals			G = 13.02	H = 11.14	J = 20.46	K = 45.10

D = 10
N = 2D = 20
SSB = $(K/2) - (J^2/N) = (45.10/2) - (20.46^2/20) = 1.62$
SST = $G + H - (J^2/N) = 13.02 + 11.14 - (20.46^2/20) = 3.23$
SSW = SST - SSB = 3.23 - 1.62 = 1.61
MSB = SSB/(D-1) = 1.62/9 = 0.18
MSW = SSW/(N-D) = 1.61/10 = 0.16
F = MSB/MSW = 0.18/0.16 = 1.13
$\bar{Y}$ = J/N = 20.46/20 = 1.02
s^2 = SST/(N-1) = 3.23/19 = 0.17

ANOVA Table

Source	Degrees of Freedom	Sum of Squares	Mean Square Error	F-statistic
Between Days	9	1.62	0.18	1.13
Within Days	10	1.61	0.16	
Total	19	3.23		

continued

Example 1 (continued)

1) Screening for Outliers

$\overline{Y} = 1.02$

$s^2_{TOT} = (MSB + MSW)/2 = (0.18 + 0.16)/2 = 0.17$

$s_{TOT} = \sqrt{s^2_{TOT}} = 0.41$

Any value outside the interval

$$(\overline{Y} - 4.0\ s_{TOT},\ \overline{Y} + 4.0\ s_{TOT}) = (-0.62,\ 2.66)$$

is considered suspect. Since the recoveries in this example range from 0.39 to 1.66, it is concluded that this data set is free of outliers.

2) Testing Day Effect

Using a 10 percent significance level ($\alpha = 0.10$), the critical value for testing the day effect is

$$F_{D-1,N-D,1-\alpha} = F_{9,10,.90} = 2.35 \qquad \text{(from Table B1-1)}$$

Since the statistic F = 1.13 (from the ANOVA table) is less than the critical value 2.35, it is concluded that the day effect is not significant.

3) Testing the System Bias

Since there is no day effect, a 95 percent CI for the true mean recovery, μ, is given by

$$Pr[\overline{Y} - t_{N-1,.975}\ \sqrt{s^2/N} \leq \mu \leq \overline{Y} + t_{N-1,.975}\ \sqrt{s^2/N}] = .95,$$

where $t_{N-1,.975}$ is the 97.5th percentile of the t-distribution with N-1 degrees of freedom. From Table B1-1, $t_{19,.975} = 2.09$.

continued

Example 1 (continued)

$$Pr[1.02 - (2.09)\ \sqrt{0.17/20} \leq \mu \leq 1.02 + (2.09)\ \sqrt{0.17/20}] = .95$$

$$Pr[0.83 \leq \mu \leq 1.21] = .95$$

For this example, suppose the EPA had specified a maximum bias $b_o = 0.10$. Therefore, a proposed method whose true mean recovery was somewhere between $(1-b_o,\ 1+b_o) = (0.90,1.10)$ would be considered accurate by the Agency's standards. Since the interval for the true mean recovery, μ, of the proposed method (0.83,1.21) overlaps the interval (0.90,1.10), it is concluded that the bias of the proposed method is within the approved limits for measuring low concentrations.[1]

4) Testing the System Precision

Since there is no day effect, a lower bound for the true variance, σ^2, of the proposed method is given by:

$$Pr[(N-1)s^2/\chi^2_{N-1,.95} \leq \sigma^2] = .95\ ,$$

where $\chi^2_{N-1,.95}$ is the 95th percentile of the chi-square distribution with N-1 degrees of freedom. From Table 4-1, $\chi^2_{19,.95} = 30.14$.

$$Pr[(19)(0.17)/30.14 \leq \sigma^2] = .95$$

$$Pr[0.11 \leq \sigma^2] = .95$$

---

[1]This test for bias has the potential to penalize labs with good precision and reward labs with poor precision. Therefore, when establishing DQOs for each petition, OSW sets precision levels accordingly to avoid this problem.

continued

Example 1 (continued)

For this example, suppose the EPA had specified a maximum variance of $\sigma_o^2 = 0.25$. Since the lower bound of the true variance (0.11) is less than or equal to $\sigma_o^2 = 0.25$, it is concluded that the precision of the proposed method meets with the Agency's standards for measuring low concentrations.

continued

Example 1 (continued)

B. High Concentration

Day Recoveries

i	Y_{1i}	Y_{2i}	Y_{1i}^2	Y_{2i}^2	$T_i = Y_{1i} + Y_{2i}$	T_i^2
1	1.31	1.21	1.72	1.46	2.52	6.35
2	1.47	1.06	2.16	1.12	2.53	6.40
3	0.86	1.31	0.74	1.72	2.17	4.71
4	1.02	0.15	1.04	0.02	1.17	1.37
5	0.23	0.42	0.05	0.18	0.65	0.42
6	0.34	1.04	0.12	1.08	1.38	1.90
7	1.00	1.69	1.00	2.86	2.69	7.24
8	1.61	1.26	2.59	1.59	2.87	8.24
9	0.93	0.93	0.86	0.86	1.86	3.46
10	0.74	0.46	0.55	0.21	1.20	1.44
Totals			G = 10.83	H = 11.10	J = 19.04	K = 41.53

D = 10
N = 2D = 20
SSB = $(K/2) - (J^2/N) = (41.53/2) - (19.04^2/20) = 2.64$
SST = $G + H - (J^2/N) = 10.83 + 11.10 - (19.04^2/20) = 3.80$
SSW = SST - SSB = 3.80 - 2.64 = 1.16
MSB = SSB/(D-1) = 2.64/9 = 0.29
MSW = SSW/(N-D) = 1.16/10 = 0.12
F = MSB/MSW = 0.29/0.12 = 2.42
$\bar{Y}$ = J/N = 19.04/20 = 0.95
s^2 = SST/(N-1) = 3.80/19 = 0.20

ANOVA Table

Source	Degrees of Freedom	Sum of Squares	Mean Square Error	F-statistic
Between Days	9	2.64	0.29	2.42
Within Days	10	1.16	0.12	
Total	19	3.80		

continued

Example 1 (continued)

1) Screening for Outliers

$\overline{Y} = 0.95$

$s^2_{TOT} = (MSB + MSW)/2 = (0.29 + 0.12)/2 = 0.21$

$s_{TOT} = \sqrt{s^2_{TOT}} = 0.46$

$(\overline{Y} - 4.0\ s_{TOT}, \overline{Y} = 4.0\ s_{TOT}) = (-0.89, 2.79)$

Since all recoveries fall within this interval, it is concluded that this data set is free of outliers.

2) Testing Day Effect

Since the statistic $F = 2.42$ (from the ANOVA table) is greater than or equal to the critical value $F_{9,10,.90} = 2.35$, it is concluded that a significant day effect is present.

3) Testing the System Bias

Since there is a day effect, a 95 percent CI for the true mean recovery, μ, is given by

$$Pr[\overline{Y} - t_{D-1,.975}\sqrt{MSB/N} \leq \mu \leq \overline{Y} + t_{D-1,.975}\sqrt{MSB/N}] = .95$$

where $t_{D-1,.975}$ is the 97.5th percentile of the t-distribution with D-1 degrees of freedom. From Table B1-1, $t_{9,.975} = 2.26$.

$$Pr[0.95 - 2.26\sqrt{0.29/20} \leq \mu \leq 0.95 + 2.26\sqrt{0.29/20} = .95$$

$$Pr[0.68 \leq \mu \leq 1.22] = .95$$

continued

Example 1 (continued)

Since this interval overlaps the limits set by the EPA for mean recovery $(1-b_o, 1+b_o) = (.90,1.10)$, it is concluded that the bias of the proposed method is considered acceptable by the Agency's standards.

4) Testing the System Precision

Since there is a day effect, the following equations must be calculated:

$$y = (MSB + MSW)/2 = (.29 + .12)/2 = 0.21$$

$$n = \frac{4y^2}{(MSB^2/D-1) + (MSW^2/N-D)} = \frac{4(.21)^2}{(.29^2/9) + (.12^2/10)} = 16.36$$

$\therefore$ n = 17 (rounding n to next highest integer).

A 95 percent CI for the lower bound of the true variance, σ^2, of the proposed method is

$$Pr[ng/\chi^2_{n,.95} \leq \sigma^2] = .95 ,$$

where $\chi^2_{n,.95}$ is the 95th percentile of the chi-square distribution with n degrees of freedom. From Table B1-1, $\chi^2_{17,.975} = 27.59$.

$$Pr[(17)(0.21)/27.59 \leq \sigma^2] = .95$$

$$Pr[0.13 \leq \sigma^2] = .95$$

continued

Example 1 (continued)

Since this lower bound is less than or equal to the maximum variance specified by the EPA, $\sigma^2 = 0.25$, the precision of the proposed method meets with the Agency's standards for measuring high concentrations.

Conclusion: Two hypotheses were tested for the proposed method:

$$H_1: \quad |b_{new}| \leq b_o$$

$$H_2: \quad \sigma^2_{new} \leq \sigma^2_o,$$

where b_o and σ^2_o are data quality objectives specified by the EPA. The recovery data obtained from this method were analyzed separately for low and high concentrations (parts A and B, respectively). In both cases, the data supported the hypotheses tested; therefore, the proposed method would be accepted by the Agency as an approved procedure for measuring solid wastes.

B1.4 Statistical Analyses for Single-Site Cases with Comparative Data Quality Objectives

This section describes statistical procedures for testing equivalency of the proposed and approved methods, with respect to precison and bias. It is assumed that preliminary analyses to test for outliers and equality of replicate variances (described in Section B1.2) have already been carried out. Under this approach, the approved method is considered the standard (i.e., its measurements are assumed to be "correct"), and the proposed method is compared with it. The test hypotheses, H_3 and H_4, specified in Section 4.2, reflect similar precision and bias for the two methods. If either of these hypotheses is rejected, then the proposed method fails the equivalency test. For this reason, the comparative approach is not recommended when the petitioner suspects that his proposed method is more accurate than (and therefore not equal to) the approved method. As noted in Section 4.2, if there is reason to believe that the proposed method is less biased or more precise than the approved method, then the petitioner should request using absolute rather than comparative data quality objectives. It is assumed that two replicate measurements are taken for each sample and that any modification on the minimal design is in the direction of additional days/sample. Table B1-1 can be used to obtain appropriate percentiles for the statistical tests when the design has been modified as such. An example is provided at the end of this section to clarify the analysis discussed here.

The first step in this analysis is to complete an ANOVA table for each method. The layout of the data and the necessary formulas are the same as for the absolute case, with method substituted for concentration (see Table B1-2). It is not necessary to calculate the F-statistics for this analysis. Once the ANOVA tables have been completed, the petitioner may use the between-day and

within-day mean square errors (MSB and MSW) to check for outliers in each method as described in Secton B1.2.1. Note that once an outlier has been found and treated, then all results affected by it must be recalculated.

The precision of the proposed method can be compared to that of the approved method by obtaining a 95 percent confidence interval for the ratio of the true replicate variances, σ_p^2 (proposed method) and σ_a^2 (approved method),

$$\Pr\left(\frac{MSW_p}{MSW_a(F_{N-D,\ N-D,.975})} \leq \frac{\sigma_p^2}{\sigma_a^2} \leq \frac{MSW_p}{MSW_a(F_{N-D,N-D,.025})}\right) = .95$$

where MSW_p and MSW_a are the within-day mean square errors (from the ANOVA tables) for the proposed and approved methods, respectively, and $F_{N-D,N-D,.975}$ and $F_{N-D,N-D,.025}$ are the 97.5th and 2.5th percentiles of the F-distribution with (N-D) degrees of freedom in both the numerator and denominator. Note that

$$F_{N-D,N-D,.025} = \frac{1}{F_{N-D,N-D,.975}} .$$

For the minimal design with 10 days (N=20), $F_{10,10,.975} = 3.72$ and $F_{10,10,.025} = 0.27$, and the above CI simplifies to

$$\Pr\ \frac{MSW_p}{MSW_a(3.72)} \leq \frac{\sigma_p^2}{\sigma_a^2} \leq \frac{MSW_p}{MSW_a(0.27)} = .95$$

$$\Pr\ \left(0.27\frac{MSW_p}{MSW_a} \leq \frac{\sigma_p^2}{\sigma_a^2} \leq \frac{MSW_p}{MSW_a(0.27)}\right) = .95$$

Substituting MSW for the two methods yields upper and lower bounds for the ratio of the true variances. If this interval contains the value 1, corresponding to $\sigma_p^2 = \sigma_a^2$, then equivalency of the method precisions is supported. In contrast, an interval that does not contain the value 1 implies that the

proposed and approved methods differ with respect to precision, and the proposed method would not be approved.

If the method precisions are considered equal, then the petitioner can proceed to compare their bias. This is done by performing a two-way analysis of variance with method and days as the factors. A method by day interaction is also considered in this analysis. The calculations for the two-way ANOVA (shown in Table B1-3) are easily obtained from those of the one-way ANOVAs already performed for each method. Note that the subscripts "a" and "p" are used in Table B1-3 to indicate totals from the one-way ANOVA calculations for the approved and proposed methods, respectively. Once the two-way ANOVA table has been completed, the method by day interaction is tested using the F-statistic

$$F = \frac{MSMD}{MSE}.$$

If this value is greater than or equal to the critical value $F_{D-1,2D,.95}$, then a significant interaction is concluded. For the minimal design, $F_{D-1,2D,.95} = F_{9,20,.95} = 2.39$. A method by day interaction indicates that the difference between the two methods is not the same for all days. Therefore, if the method by day interaction is statistically significant, the proposed method will not be accepted, as there is sufficient evidence that the methods differ for some of the days. Plots of the data may help to examine the sources of interaction more closely.

If the method by day interaction is not significant, its sum of squares (SSMD) and degrees of freedom (D-1) are added to those for the error to form a new mean square error:

$$MSE^* = \frac{SSE + SSMD}{2D + (D-1)} = \frac{SSE + SSMD}{3D - 1}.$$

Table B1-3. TWO-WAY ANALYSIS OF VARIANCE (ANOVA) TABLE

Source	Degrees of Freedom	Sum of Squares	Mean Square Error
Method	1	SSM	MSM
Day	(D-1)	SSD	MSD
Method x Day	(D-1)	SSMD	MSMD
Error	2D	SSE	MSE
Total	4D-1	SST	

$C = (J_a + J_p)^2/4D$

$SST = (G_a + H_a + G_p + H_p) - C$

$SSM = [(J_a^2 + J_p^2)/2D] - C$

$SSD = [\sum_{i=1}^{D} (T_{ia} + T_{ip})^2]/4 - C$

$SS\ (replicates) = [(K_a + K_p)/2] - C$

$SSMD = SS\ (replicates) - SSD - SSM$

$SSE = SST - SSD - SSM - SSMD$

$MSM = SSM/1 = SSM$

$MSD = SSD/(D-1)$

$MSMD = SSMD/(D-1)$

$MSE = SSE/2D$

NOTE: The totals G, H, J, K, and T_i are calculated in the one-way ANOVA tables for each method (see Table B1-2). The subscripts "a" and "p" denote the approved and proposed test methods.

The main effect for method can then be tested with the F-statistic

$$F = \frac{MSM}{MSE^*} .$$

If this value is greater than or equal, the critical value $F_{1,3D-1,.95}$ (= 4.18 for minimal design), then a significant method effect is concluded (i.e., the methods do not yield equivalent measurements). Therefore, the proposed method will not be accepted.

Numerical Example 2

Hypothetical measurements from waste samples tested by two methods (proposed and approved) are analyzed to illustrate some of the analytical procedures just described.

Example 2. Analysis of Measurements from The Proposed And Aproved Methods for the Single Site Case with Comparative DQOs

A. Proposed Method: one-way ANOVA

Day	Logarithms of Measurements					
i	Y_{1i}	Y_{2i}	Y_{1i}^2	Y_{2i}^2	$T_i = Y_{1i} + Y_{2i}$	T_i^2
1	2.15	2.17	4.62	4.71	4.32	18.66
2	2.15	2.30	4.62	5.29	4.45	19.80
3	2.02	1.84	4.08	3.39	3.86	14.90
4	2.03	1.78	4.12	3.17	3.81	14.52
5	1.91	1.81	3.65	3.28	3.72	13.84
6	2.19	1.82	4.80	3.31	4.01	16.08
7	2.08	2.27	4.33	5.15	4.35	18.92
8	2.30	2.01	5.29	4.04	4.31	18.58
9	2.06	2.10	4.24	4.41	4.16	17.31
10	2.10	2.16	4.41	4.67	4.26	18.15
Totals			G_p= 44.16	H_p= 41.42	J_p= 41.25	K_p= 170.76

D = 10
N = 2D = 20
SSB = $(K/2) - (J^2/N) = (170.76/2) - (41.25^2/20) = 0.302$
SST = $G + H - (J^2/N) = 44.16 + 41.42 - (41.25^2/20) = 0.502$
SSW = SST - SSB = 0.502 - 0.302 = 0.200
MSB = SSB/(D-1) = 0.302/9 = 0.034
MSW = SSW/(N-D) = 0.200/10 = 0.020
$\overline{Y}_2$ = J/N = 41.25/20 = 2.063
s^2 = SST/(N-1) = 0.502/19 = 0.026

ANOVA Table

Source	Degrees of Freedom	Sum of Squares	Mean Square Error
Between Days	9	0.302	0.034
Within Days	10	0.200	0.020
Total	19	0.502	

continued

Example 2 (continued)

B. Approved Method: one-way ANOVA

Day	Logarithms of Measurements					
i	Y_{1i}	Y_{2i}	Y_{1i}^2	Y_{2i}^2	$T_i = Y_{1i} + Y_{2i}$	T_i^2
1	2.32	2.30	5.38	5.29	4.62	21.34
2	2.36	2.27	5.57	5.15	4.63	21.44
3	2.18	2.29	4.75	5.24	4.47	19.98
4	2.18	1.92	4.75	3.69	4.10	16.81
5	1.94	2.00	3.76	4.00	3.94	15.52
6	2.01	2.21	4.04	4.88	4.22	17.81
7	2.24	2.40	5.02	5.76	4.64	21.53
8	2.39	2.32	5.71	5.38	4.71	22.18
9	2.21	2.21	4.88	4.88	4.42	19.54
10	2.11	2.03	4.45	4.12	4.14	17.14
Totals			$G_a = 48.31$	$H_a = 48.39$	$J_a = 43.89$	$K = 193.29$

$D = 10$
$N = 2D = 20$
$SSB = (K/2) - J^2/N) = (193.29/2) - (43.89^2/20) = 0.328$
$SST = G + H - (J^2/N) = 48.31 + 48.39 - (43.89^2/20) = 0.383$
$SSW = SST - SSB = 0.383 - 0.328 = 0.055$
$MSB = SSB/(D-1) = 0.328/9 = 0.036$
$MSW = SSW/(N-D) = 0.055/10 = 0.006$
$\overline{Y} = J/N = 43.89/20 = 2.195$
$s^2 = SST/(N-1) = 0.383/19 = 0.020$

ANOVA Table

Source	Degrees of Freedom	Sum of Squares	Mean Square Error
Between Days	9	0.328	0.036
Within Days	10	0.055	0.006
Total	19	0.383	

continued

Example 2 (continued)

C. Two-Way ANOVA

$C = (J_a + J_p)^2/4D = (43.89 + 41.25)^2/(4)(10) = 181.22$

$SST = G_a + H_a + G_p + H_p - C = 48.31 + 48.39 + 44.16 + 41.42 - 181.22 = 1.060$

$SSM = [(J_a^2 + J_p^2)/2D] - C = [(43.89^2 + 41.25^2)/(2)(10)] - 181.22 = 0.175$

$SSD = [\sum_{i=1}^{D} (T_{ia} + T_{ip})^2]/4 - C = [(727.09)/4] - 181.22 = 0.553$

$SS\ (replicates) = [(K_a + K_p)/2] - C = [(193.29 + 170.76)/2] - 181.22 = 0.805$

$SSMD = SS\ (replicates) - SSD - SSM = 0.805 - 0.553 - 0.175 = 0.077$

$SSE = SST - SSD - SSM - SSMD = 1.060 - 0.553 - 0.175 - 0.077 = 0.255$

$MSM = SSM/1 = 0.175$

$MSD = SSD/(D-1) = 0.553/9 = 0.061$

$MSMD = SSMD/(D-1) = 0.077/9 = 0.009$

$MSE = SSE/2D = 0.255/(2)(10) = 0.013$

ANOVA Table

Source	Degrees of Freedom	Sum of Squares	Mean Square Error
Method	1	0.175	0.175
Day	9	0.553	0.061
Method x Day	9	0.077	0.009
Error	20	0.255	0.013
Total	39	1.060	

continued

Example 2 (continued)

1. Screening for Outliers

Proposed Method: $\overline{Y} = 2.063$

$$s^2_{TOT} = (MSB + MSW)/2 = (0.034 + 0.020)/2 = 0.027$$

$$(\overline{Y} \pm 4.0 \sqrt{s^2_{TOT}}) = (2.063 \pm 4.0 \sqrt{.027}) = 1.406, 2.720)$$

Approved Method: $\overline{Y} = 2.195$

$$s^2_{TOT} = (MSB + MSW)/2 = (0.036 + 0.006)/2 = 0.021$$

$$(\overline{Y} \pm 4.0 \sqrt{s^2_{TOT}}) = (2.195 \pm 4.0 \sqrt{.021}) = 1.615, 2.775)$$

Since the observations for both methods fall within the specified intervals, it is concluded that this data set is free of outliers.

2. Testing Equality of the Method Replicate Precisions

A 95% CI for the ratio of the true replicate variances σ^2_p (proposed method) and σ^2_a (approved method) is given by:

$$Pr\left(\frac{MSW_p}{MSW_a(F_{N-D,N-D,.975})} \leq \frac{\sigma^2_p}{\sigma^2_a} \leq \frac{MSW_p}{MSW_a(F_{N-D,N-D,.025})} \right) = .95$$

where $F_{N-D,N-D,.975}$ is the 97.5th percentile of the F-distribution with (N-D) degrees of freedom in both the numerator and denominator, and $F_{N-D,N-D,.025} = \frac{1}{F_{N-D,N-D,.975}}$.

From Table B1-1, $F_{9,9,.975} = 3.72$ and $F_{9,9,.025} = \frac{1}{3.72} = 0.27$.

$$Pr\left(\frac{0.020}{(0.006)(3.72)} \leq \frac{\sigma^2_p}{\sigma^2_a} \leq \frac{0.020}{(0.006)(0.27)} \right) = .95$$

$$Pr\left(0.90 \leq \frac{\sigma^2_p}{\sigma^2_a} \leq 12.35\right) = .95$$

continued

Example 2 (continued)

Since the above interval includes the value 1, corresponding to $\sigma_p^2 = \sigma_a^2$, it is concluded that the method replicate precisions are not significantly different.

3. Testing the Method x Day Interaction

Using a 5 percent significance level ($\alpha = 0.05$), the critical value for testing the interaction term is

$$F_{D-1,2D,1-\alpha} = F_{9,20,.95} = 2.39 \qquad \text{(from Table B1-1)}$$

Since $F = \frac{MSMD}{MSE} = \frac{0.009}{0.013} = 0.692$ is less than the critical value 2.39, it is concluded that the interaction term is not significant.

4. Testing the Method Main Effect

Since the interaction term was not significant, its sum of squares (SSMD) and degrees of freedom (D-1) can be added to those for the error, and a new mean square error (MSE^*) can be calculated.

$$MSE^* = \frac{SSE + SSMD}{2D + (D-1)} = \frac{0.255 + 0.077}{(2)(10) + (9)} = 0.011$$

The critical value for testing the main method effect is $F_{1,3D-1,1-\alpha} = F_{1,29,.95} = 4.18$ (from Table B1-1). Since $F = \frac{MSMD}{MSE^*} = \frac{0.175}{0.011} = 15.91$ is greater than or equal to the critical value 4.18, it is concluded that the two methods are not equivalent, and, therefore, the proposed method would not be accepted by the EPA.

continued

Example 2 (continued)

Conclusion: Although the method replicate precisions were determined to be similar, the average measurement for the proposed method did not concur with that for the approved method, as indicated by the method main effect. Therefore, the proposed method would not be accepted by the EPA.

B2. EXPERIMENTAL DESIGNS AND METHODS OF DATA ANALYSIS FOR MULTISITE CASES

If cost permitted, extensive experiments involving many sites, samples, and laboratories could be done to estimate bias, individual variance components and total measurement system variance for the range of conditions over which the method would be applied. Unfortunately such experiments tend to be lengthy and economically unfeasible. A more compact, specially designed experiment may assure that relevant sources of test method variance are built in (in the absolute case) or controlled for (in the comparative case), while reducing the study to a more manageable period of time, such as one or two weeks. The designs suggested in this document are intended to incorporate realistic total system variance rather than to resolve individual components thereof.

B2.1 Experimental Design Terminology

Experiments are designed to investigate the effects of certain factors on variables of interest, dependent variables, or response variables (y). In this document y usually denotes either (1) waste sample recovery (measurement/concentration) in the absolute case, or (2) logarithm of measurement of a given analyte in the comparative case.

A factor is a variable, such as a waste site or method of chemical analysis. A factor is divided into categories for the experiment called levels of the factor, e.g., GC/MS, variations 1 and 2.

The influence of factors on outcome or response are called effects.

The mathematical description of an experiment to investigate the effects of n factors on the variable of interest is called an n-way classification.

A fixed effect such as a quantitative analysis method (treatment) is assumed to be constant. Random effects include laboratory or analyst where the lab or analyst is regarded as a randomly selected member of a population.

Further discussion of the difference can be found in Winer (3). Combinations of fixed and random effects are called mixed effects.

The experimental situations for equivalency testing fall largely outside fixed effects analysis of variance, where interest centers mainly on the estimation of functions of means. Although the effects of quantitative analytical methods are considered fixed (nonrandom), effects due to sampling within sites are usually considered random, as are effects due to differences in day-to-day setups of equipment and preparations. Laboratory effects are often considered random, unless, for example, only a small, select group of laboratories can perform a test method. For such random effects, variances (not means) are the statistical measure of primary interest. Inference about linear combinations of variances leads to complicated formulas about their variances (variances of variances). The designs proposed here then involve mixed effects.

A crossed (2-way) classification combines every level of one factor with every level of another factor. For instance, each of a given number of waste samples might be split into four parts and tested with duplicated analyses by each of two test method procedures; in this case, the samples are crossed with method of analysis, as in the simple single-site comparative case of Section 4.

Nested classifications preclude the formation of all possible combinations, by pairing each level of one factor with only one level (rather than all levels) of the other factor. An example is waste samples nested within a waste site, where samples are unique to particular hazardous waste site (Figure B2-1). With more than two factors, hybrids of nested and crossed classifications can occur. Table B2-1 lists and describes several pertinent factors that might be involved in the experimental designs.

a) A Crossed Classification: Test Method Crossed With Waste Sample

Sample \ Method	1	2
1	$Y_{1,1}$	$Y_{1,2}$
2	$Y_{2,1}$	$Y_{2,2}$

b) A Nested Classification: Waste Sample Nested Within Waste Site

Site 1		Site 2	
Sample 1	Sample 2	Sample 3	Sample 4
$Y_{1,1}$	$Y_{1,2}$	$Y_{2,3}$	$Y_{2,4}$

Figure B2-1. Examples of crossed and nested classifications.

Legend. Y = test method's recovery of the hazardous waste property or constituent of interest. In practice, samples 3 and 4 are often referred to as samples 1 and 2, i.e., the same labels are often used for samples from both sites, even though physically distinct sets of samples are involved. In this case, $Y_{2,3}$ and $Y_{2,4}$ become $Y_{2,1}$ and $Y_{2,2}$.

TABLE B2-1. SOME COMMON FACTORS IN EXPERIMENTAL DESIGNS FOR EQUIVALENCY TESTING

Factor	Fixed or Random Effect	Crossed or Nested	Index	Number of Levels
Method of Chemical Analysis	Fixed	Crossing factor	m	M
Hazardous Waste Site	Random or fixed	Crossing factor	t	T
Day	Random	Crossing factor	d	D
Sample	Random	Nested within site	s	S
Concentration	Fixed	Crossing factor	c	C
Laboratory	Random or fixed	Crossing factor	l	L
Replicate	Random	Nested within cell	r	R

Legend. The indices are used as subscripts for model terms and to designate factors and their combinations (interactions). The number of levels is the upper limit on the index. For instance, for a comparative study there are usually two methods so m can be 1 or 2, i.e., m = 1, 2. For a study with five sites T = 5 and t runs from 1 to 5, t = 1, ..., 5. Notation such as Y_{mtcr} is used to denote the measurement using method m the rth replicate of a sample at concentration c from site t.

B2.2 Classification of Factors as Random or Fixed

Factors such as laboratory and site are usually regarded as random in order that results may be generalized to arbitrary laboratories and sites rather than limited to those actually employed in the study. If the petitioner does not mind the conclusion being restricted to those sites and laboratories used in the study, then these may be treated as fixed effects. If the laboratories in the study are the only ones capable of using the method, then laboratory must be regarded as fixed, since there is no larger reference population. If the sites were selected out of convenience, or are the only sites where the method is proposed for use, then site may be treated as a fixed effect.

It is anticipated that most multisite and multilaboratory petitions will be from vendors of equipment and professional associations, who will want to treat these effects as random. The statistical analyses suggested here are for this case, i.e., treat laboratory and site as random, while treating method and concentration as fixed.

The absolute and comparative cases are as defined in Section 4. The example experimental designs offered here are three-way factorial designs. The factors are day, site, and concentration in the absolute case; and day, site, and method in the comparative case. In both cases day and site are regarded as random and the third factor is regarded as fixed. The analyses of variance and tables of expected mean squares for other three-way factorials (such as all factors fixed or two fixed, one random) can be found in Kleinbaum and Kupper (4, p. 367). For multilaboratory petitions, the design can be obtained by replacing day with laboratory in either the absolute or comparative case. Each laboratory should then analyze samples from different

sites on different days, and the laboratories should not discuss their results or otherwise collude.

B2.3 Multisite Absolute Cases

An important consideration in experimental design for equivalency testing is whether waste site is regarded as a fixed or random effect. If the test method is proposed for use at a small, select group of waste sites - most or all of which are sampled for the equivalency data - then site is a fixed effect. However, if the test method is planned for general use at any site, then site should probably be modeled as a random effect. This means the waste sites should be selected at random. The statistical treatment in the following guidance assumes site is a random effect.

Only one type of design is discussed for multisite petition cases. The petitioner may propose others if financial, logistic, or other considerations dictate. At least five sites should be selected at random. The example of Figure B2-2 involves taking thirty samples at random from within each site and spiking and analyzing six per day (two analyses at each of three concentrations) over a period of five days. The days do not need to be consecutive. In fact, EPA usually prefers days to be a week or more apart, involving separate independent setups and preparations. The recommended spiking concentrations are T/2, T, 3T/2 where T is the regulatory threshold for the analyte.

B2.4 Comparative Data Quality Objectives, Multisite Case

This approach compares the proposed test method directly with the approved test method. Waste samples are not spiked in this approach; the dependent variable is the test method measurement (or logarithm thereof)

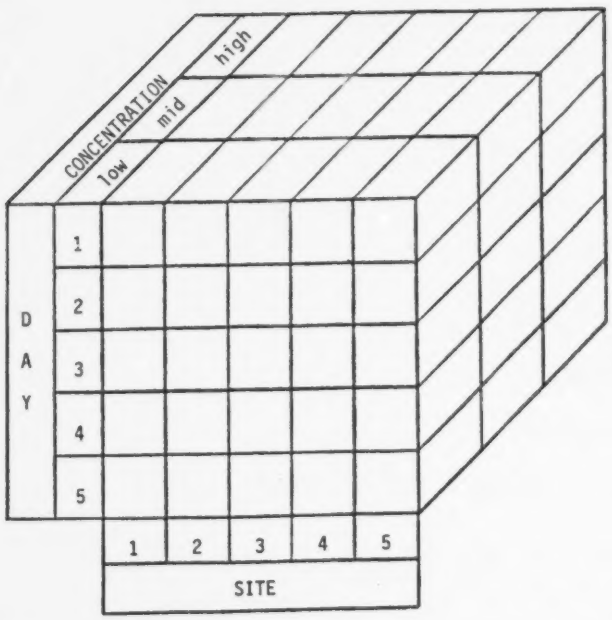


Figure B2-2. Experimental design for multisite absolute case: concentration (fixed effect) crossed with day and site (random effects).

Expected Mean Squares for Multisite Absolute Case (3, p. 203)

Source of Variation	d.f.	Expected Mean Squares
c (concentration)	C-1	V+RV(csd)+RSV(cd)+RDV(cs)+RSDV(c)
s (site)	S-1	V+CRV(sd)+CRDV(s)
d (day)	D-1	V+CRV(sd)+CRSV(d)
cs	(C-1)(S-1)	V+RV(csd)+RDV(cs)
cd	(C-1)(D-1)	V+RV(csd)+RSV(cd)
sd	(S-1)(D-1)	V+CRV(sd)
csd	(C-1)(S-1)(D-1)	V+RV(csd)
within-cell (error)	CSD(R-1)	V

Figure B2-2 (continued)

Estimates of Variance Components

$\hat{v}$ = MS(error)

$\hat{v}$ (csd) = [MS(csd) - MS(error)]/R

$\hat{v}$ (sd) = [MS(sd) - MS(error)]/CR

$\hat{v}$ (cd) = [MS(cd) - MS(csd)]/RS

$\hat{v}$ (cs) = [MS(cs) - MS(csd)]/RD

$\hat{v}$ (d) = [MS(d) - MS(sd)]/CRS

$\hat{v}$ (s) = [MS(s) - MS(sd)]/CRD

$\hat{v}$ (c) = [MS(c) - MS(cd) - MS(cs) + MS(csd)]/RSD

Figure B2-2 Legend: Two replicate observations per cell. Thirty samples are randomly selected from each of five randomly selected sites, with six samples per site analyzed per day (two at each concentration). For multilaboratory petitions replace day by laboratory. Each laboratory should then analyze samples from different sites on different days.

Outline of Analysis for Multisite Absolute Case

[Additional details are in Section B3]

1. Screen for Outliers [Section B3.2]

2. Check for equality of replicate variances of recoveries; transform if necessary [Section B3.3]

3. Prior to testing adequacy of bias and total system variance, test for model simplification, beginning with third order interaction, then second order interaction, then main effects.

   Pooling of sum of squares and deleting nonsignificant variance components for remaining tests is recommended. In addition to presumably increasing power of tests, this can lead to an exact test of the main fixed concentration effect (an exact test does not exist for the full model; see the table of expected mean squares in Figure B2-2). Pooling is done by combining sums of squares for nonsignificant terms with the error sum of squares and also adding the corresponding degrees of freedom to those for error (Winer, 1962, p. 202). The new mean squared error is the ratio of the new sum of squares and the new degrees of freedom. The significance level for preliminary tests of these secondary hypotheses should be at a higher level (e.g., 20% to 30%) than tests of primary hypotheses.

4. Model validation: Perform model diagnostic tests on residuals for the simplified model, including a test of normality. (5)

Figure B2-2 (continued)

5. Test adequacy of total system variance: Using all model terms which remain after the preliminary tests, i.e., were judged to be significant, estimate the total system variance by estimating and summing the corresponding variance components. Use Satterthwaite's approximation (6, p. 369) to obtain a 95% lower confidence limit for total system variance. If this lower limit is greater than σ^2, the adequacy of precision is rejected. In some cases, the improved approximation of Welch should be used (6, p. 370). Additional details are in Section B3.

6. Test method unbiasedness: In general, bias is carried by those model terms involving only fixed effects. In the present 3-way mixed model this is the grand mean plus the main concentration effect; therefore, first for a main concentration effect. Usually the second order (cd) interaction will be nonsignificant from step 3, so that an exact test of a concentration effect can be based on

$$F = MS(c)/MS(cs).$$

If no exact test exists, then a test can be based on quasi-F ratios (Winer 1962, p. 199). If a difference among the concentrations exists, then each concentration must be tested separately for adequacy of bias. Multiple comparisons procedure (6) may be used to control the overall error rate. If no difference exists, test the grand mean μ i.e., test H: $|\mu - 1| < b_0$.

rather than the method's recovery (measurement/true value). If the petitioner believes that the proposed method is less biased than the approved method, then absolute criteria should be used to assess equivalency.

If there are any common elements to the two test methods, then it is best to apply them concurrently (e.g., the same day) to a given waste sample. This proposed experimental design is a type of paired comparison, designed to eliminate by blocking effects which are not of primary interest. Each waste sample is split into four parts and two subsamples are tested by each method, thus controlling for sample effects by testing each sample by both methods. A waste sample should be tested concurrently with the two methods to equalize sample holding times.

Since it is the objective of comparative DQOs to compare the proposed and approved methods directly, waste samples are subsampled or split into two parts to control for differences among samples. To approximate the range of measurements encountered once a test method is in practice, one varies the types of waste samples tested by selecting several samples at random from several randomly selected sites. The randomization at all levels of measurement is necessary if the petitioner wants to generalize the results to arbitrary waste sites. The petitioner must be careful to obtain representative subsamples or splits of waste. As a further precaution, the two subsamples should be randomized to the two test methods. Figure B2-3 contains an example experimental design with five waste sites and five days and an outline of suggested analyses. Additional details of analysis are in Section B3.

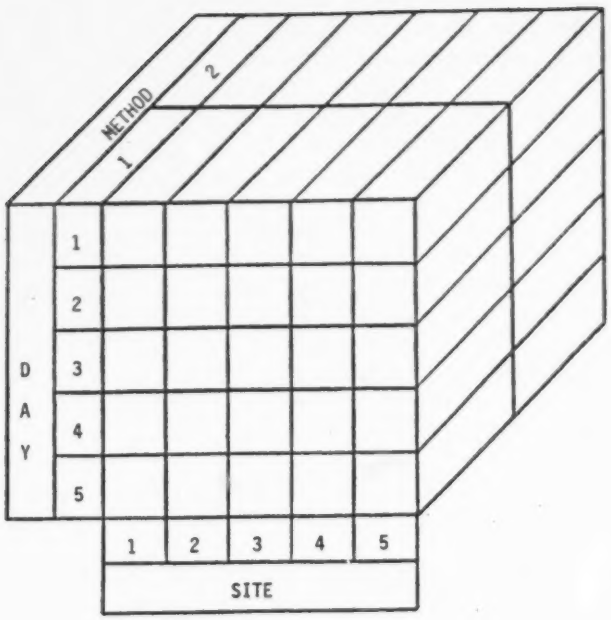


Figure B2-3. Experimental design for multisite comparative case: method (fixed effect) crossed with day and site (random effects)

Expect Mean Squares for Multisite Comparative Case (3, page 203)

Source of Variation	degrees of freedom	Expected Mean Squares
m (method)	1	V+RV(msd)+RSV(md)+RDV(ms)+RSDV(m)
s (site)	S-1	V+2RV(sd)+2RDV(s)
d (day)	D-1	V+2RV(sd)+2RSV(d)
ms	S-1	V+RV(msd)+RDV(ms)
md	D-1	V+RV(msd)+RSV(md)
sd	(S-1)(D-1)	V + 2RV(sd)
msd	(S-1)(D-1)	V+RV(msd)
within-cell	SD(R-1)	V

Figure B2-3 (continued)

Legend: Two replicate observations per cell are recommended. Five samples are randomly selected from each of five randomly selected sites with four representative subsamples extracted from each sample. On each of five days, one sample from each site is analyzed, two subsamples by each method. The design over sites and days hopefully incorporates realistic spread of experimental conditions representative of the proposed use of the method. Within each condition (site x day cell) for each method, replicate variance is calculated and these are averaged over the cells. The average within-method replicate variances are compared by a standard F-test. The comparison of the two methods is completed by exploiting the pairing. For comparative studies, the criterion is usually that the methods be essentially equivalent, so that each model term involving method should be nonsignificant.

Outline of Analysis for Multisite Comparative Case

1. Screen for outliers [Section B3.2].

2. Check for equality of replicate variances of measurements or their logarithms; transform if necessary [Section B3.3].

3. Precision check: Compare replicate variances for the two methods. The replicate variance estimates are the average within-cell variances, averaged over the S x D cells, with degrees of freedom DS(R-1). These can be obtained as the residual mean squared errors from fitting full (i.e., include both main effects, site and day, and interaction) two-way ANOVA models separately to each method. The null hypothesis is

$$H: \quad \sigma_1^2 / \sigma_2^2 \leq 1$$

where "1" indexes the proposed method and "2" indexes the approved method. The test statistic is

$$F = \hat{\sigma}_1^2 / \hat{\sigma}_2^2$$

with DS(R-1), DS(R-1) degrees of freedom.

4. Other tests: Every model term involving method should be nonsignificant in order for the methods to be equivalent. Starting with the full 3rd order model, test the significance of third order interaction, then site * day interaction, then second order interactions involving method, then the main method effect. Pooling of sum of squares and deleting (setting equal to zero) nonsignificant variance components for remaining tests is recommended. In addition to presumably increasing power of tests, this can lead to an exact test of the main method effect (an exact test does not exist for the full model). Pooling is done by combining (adding) sums of squares for nonsignificant terms with the error sum of squares and also adding the corresponding degrees of freedom to those for error (3, p. 202). The new mean squared error is just the ratio of the new sum of squares and the new degrees of freedom. If at any stage a term involving method is significant, the testing can stop with the conclusion that a method related effect exists.

5. Model validation: Perform diagnostic tests on residuals for the final model including a normality test on residuals (5).

B2.5 Petitions Based on Quality Control Data

It is possible to characterize a test method partially from evaluation of relevant QC data taken from the method's routine day-to-day use, provided that the waste samples tested are comparable to those wastes being petitioned for testing. However, it must be acknowledged that QC test data can be fraught with complexities, due to the likely effects of site, waste sample, lab technician, and so forth. The QC data will, in most cases, not exhibit the symmetries and sample size balances possible in a specifically designed experiment. The QC data experimental design may change from day-to-day with the number of routine analyses. Also, site effects can take very complicated forms, e.g., all tests performed one week could involve only one site, while the following week's tests might involve three sites. The QC program might call for spiking samples at low, medium, and high concentrations in the working range of the test method. On a slow day with only a few scheduled routine tests, perhaps only one or two QC spiked concentrations would actually be used (unless the QC program demands all three concentrations on every day of operation), resulting in empty cells on some days.

In summary, proper statistical treatment of QC data requires keeping track of site, sample, spiked concentrations and day of analysis. Sample is nested in site and probably in day, while site and spiking concentration are crossed with day. To ignore all these effects and throw the data into a one-way analysis of variance to estimate between-day and within-day components of variance is risky at best. Proper analysis of such data requires statistical expertise.

B3. STATISTICAL SUMMARIES AND ANALYSIS

This section further details steps of the statistical analyses for multisite cases which were outlined in Figures B2-2 and B2-3. The single-site cases were treated in detail in Section 4 of the main text.

In the multisite case, if results will be generalized beyond the sites and days employed for the experimental design, these two factors must be considered random. Adding a fixed factor (concentration or method) leads to a 3-way mixed model with two random and one fixed factor. The accompanying statistical analysis can be complicated and is performed best with a computerized statistical package such as SAS, applied by an experienced statistician. A pertinent discussion of this design and its analysis is given by Winer (3, pp. 202-207).

The petitioner does have design options which lead to simpler analysis but were not treated in detail here because they are regarded as economically or logistically unfeasible. If samples can be obtained from 25 randomly selected sites and analyzed on 25 separate days, then the site factor is confounded with the day factor and simpler statistics apply. In fact, if the concentrations are in an adequately controlled (e.g., narrow) range for the test method, then preliminary tests for a concentration effect in the absolute case may indicate that the three concentrations can be regarded as just replicates and the data conform to a one-way random effects ANOVA design.

B3.1 Basic Approach

The basic approach to statistical analysis of test data is to --

- screen for outliers,
- check constancy of the waste samples' replicate precision,
- transform test data (if necessary),
- fit the full statistical model,
- test for model simplifications,
- perform diagnostic checks on the simplified model,
- take appropriate remedial action if necessary, and
- use the final selected model to test hypotheses about precision and bias.

The final testing of adequacy of precision and bias may be approached as follows:

For absolute data quality criteria, the petitioner estimates total system variance by adding estimates of each component of variance that involves a random factor and is judged by preliminary tests to be significant. Satterthwaite's chi-square approximation (described below) is used to test the hypothesis of adequacy of total system variance. Next, the petitioner estimates system biases as $\mu + \alpha_C$, i.e., the means for the different concentrations. If preliminary tests indicate there is no concentration effect, then a test on μ alone is appropriate.

In the comparative case, any statistical model terms involving the test method factor should not be significantly different than zero. Also, the replication variances for the new method should not exceed that of the proposed method.

If the experimental design's test data do not support a simplified statistical model, then the computations become involved. It is highly recommended that a computerized statistical package such as SAS be used for computations, and that a statistician be directly involved with the design and responsible for the statistical analysis of the experiment.

B3.2 Approach to Outliers and Other Model Violations

Outliers present many problems in statistical analysis. Their presence can seriously disturb conventional least squares and maximum likelihood analyses such as ANOVA. One would like to delete outliers or replace them with suitable values before statistical analysis so that the analysis is undisturbed. However, it is really only relative to a reasonable model that outliers can be determined, i.e., they are best determined after the model is fit. To break this circle, it is proposed that preliminary outlier checks based on empirical univariate and bivariate distributions (e.g., in the absolute case, recovery and recovery versus concentration, site and day) be used and that a check be applied to residuals from the final selected model, at which time normality and other diagnostic tests are also applied.

The petitioner is free to suggest other approaches such as use of Thompson's t-test or the Dixon ratio test. EPA does, however, usually prefer outlier tests which are 2-tailed at the 1 percent significance level or less. The suggested approach involves inspection of distributions of recoveries for each designed factor (from day, site, concentration, method) as well as recoveries pooled over all conditions. Any data point more than 4 standard deviations from the mean is suspect. Rather than delete an outlier observation, it is preferable to replace it with a value regarded as representative for the corresponding experimental condition, so as to preserve the balance of the design. In no case should all values for a given condition (cell) be deleted because of one outlier for the cell. In any case, the raw data must be reported with clear indication of what points were deemed outliers, how they were detected, and how they were treated.

For inputing values to replace rejected outliers, it is recommended that the additive (no interactions) model be fitted to the acceptable data and that

order statistics appropriate to that model be used. For instance, if the predicted value and estimated residual error variance from the fitted additive model are $\hat{y}$ and s, respectively, and if replicates (R) were used in the design, then to replace a high outlier use $\hat{y} + s/\sqrt{\pi}$. To replace a low outlier, use $(\hat{y} - s/\sqrt{\pi})$.

B3.3 Approach to Equal Variance Assumption

A standard assumption for ANOVA procedures is that the error variances or replication variances are equal under all conditions. Each experimental design recommended suggests two replicate measurements in order that this assumption, as well as hypotheses about higher order interactions, be testable. More than two replicates is not recommended. If the experiment can be expanded, it is generally more informative to add sites or days.

If the waste samples contain concentrations in the working range for the test method, then it will often be the case that standard deviation is proportional to concentration. If this is so, then one benefit of working with recoveries in the absolute cases and logarithms of measurements in comparative cases is that variances will be approximately equal. Therefore, it is recommended that initial statistical analyses use recoveries in the absolute case and logarithms of measurements in the comparative case. Some other transformation may be used if needed to attain normality or equal variances.

To test for equality of replicate variances, calculate the within-cell average ($\bar{y}_{isd}$) and standard deviation (S_{isd}) (of recoveries in the absolute case; of logarithms of measurements in the comparative case) for each cell. Plot log (S_{isd}) vs. log ($\bar{y}_{isd}$) and use ordinary linear regression

$$\log (S_{isd}) = a \log (\bar{y}_{isd}) + b$$

to test for dependence of variation on level. If the regression is significant, then try the power tranformation y^{1-a} (log(y) for a=1) on the raw data and test again as above for equality of variances (7, page 232).

Normality will be tested after the model has been fit and possibly simplified as part of the diagnostic tests on residuals.

B3.4 Analysis for Multisite Comparative Case

B3.4.1 Outlier Screen. For each separate test method, petitioners should study the distributions of logarithms of measurements by day, by site, and pooled over all conditions. Any datum more than 4 standard deviations from the average is a candidate for an outlier.

B3.4.2 Checks on Replicate Variances. Waste sample replicate variances are assumed to be approximately equal in each day * site cell and for each test method. For most test methods, the logarithm of measurement (or a power near zero, e.g., $y^{0.1}$) will be approximately normal and homoscedastic.

B3.4.3 Comparison of Method Replicate Precisions. The two variance estimates, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$, are the residual mean squared errors from separately fitting the two-way random effects models:

$$Y_{msdr} = \mu_m + \alpha_{ms} + \beta_{md} + \gamma_{msd} + \varepsilon_{msdr}.$$

With no missing data, the degrees of freedom for each is DS(R-1). The test statistic is--

$$F = \hat{\sigma}_1^2/\hat{\sigma}_2^2 \text{ with } DS(R-1), DS(R-1) \text{ degrees of freedom.}$$

The null hypothesis tested is--

$$H: \sigma_1^2/\sigma_2^2 \leq 1$$

where "1" indexes the proposed test method, and "2" indexes the approved test method. If this hypothesis is rejected, no further testing is necessary.

B3.4.4 Comparison of Method Biases. An ANOVA table for this experiment is in Figure B2-3. Each statistical model term involving method should be nonsignificant. Otherwise, some method-related effect appears to exist.

By sequentially testing for 3rd order interaction, then 2nd order interactions, then main effects, it is possible to obtain exact tests for each hypothesis. (No exact test of the main method effect exists in the full model.) That is, if a term involving method is significant, the testing can stop, with the conclusion that a method-related effect exists. Otherwise the testing continues with the nonsignificant interactions neglected (i.e., assumed equal to zero) in lower order terms which originally contained them. The sums of squares for nonsignificant interactions can be pooled with the error sums of squares after each stage.

The following sequence can be used. Here pooling is used only before the test of the main method effect.

To test significance of 3rd order interaction, use F = MS(MSD)/MS(error), with (S-1)(D-1), SD(R-1) degrees of freedom. In most cases, this 3rd order term will be not significant and the associated σ^2_{MSD} may be crossed out from the expected mean squares for other sources of variation. If this 3rd order interaction term is significant, then further testing is not necessary.

Petitioners should statistically test 2nd order interactions involving method by using--

F_1 = MS(MS)/MS(MSD) with S-1, and (S-1)(D-1) degrees of freedom

F_2 = MS(MD)/MS(MSD) with D-1, and (S-1)(D-1) degrees of freedom.

If any 2nd and 3rd order interactions involving method are significant, then a method-related effect is assumed to exist. If these 2nd and 3rd order interaction tests are not significant, then the main method effect remains to be tested. A pooling procedure is recommended for this to increase the power of the test potentially.

The petitioner should test the main method effect using $F = MS(M)/MS(error,pooled)$, with 1, $2SD(R-1) + SD - 1$ degrees of freedom.

B3.4.5 Diagnostic Tests. Separately for each method and combining data from the two methods, petitioners should plot residuals versus day and versus site and pooled over all conditions. Also, petitioners should plot residuals versus fitted values, then inspect for trends and outliers. A test for normality such as the Shapiro-Wilks test should be used on the pooled residuals. These graphic displays and diagnostic tests should be included in the petition to EPA.

B3.4.6 Conclusions. If the replicate variance of the new test method is significantly larger than that of the approved test method, or if any significant method-related effect is found, then the new test method is rejected.

B3.5 Analysis For Multisite Absolute Case

B3.5.1 Outlier Screen. Petitioners should look at the distributions of recoveries by day, by site, by concentration, and pooled over all conditions. Any datum more than 4 standard deviations from the average is a candidate for an outlier.

B3.5.2 Checks on Replicate Variances. Waste sample replicate variances are assumed to be approximately equal in each day * site * concentration cell and for each test method. For most test methods, the recoveries or their logarithms or a power near zero, e.g., $y^{0.1}$ will be approximately normal and homoscedastic.

B3.5.3 Preliminary Tests of Simplifying Hypotheses. These can be constructed from the table of expected mean squares, proceeding from bottom to top to form F-ratios, e.g., third order interaction is tested by F = MS(csd)/MS(within-cell), with (C-1)(S-1)(D-1) and CSD(R-1) degrees of freedom.

B3.5.4 Testing Adequacy of Total System Variance. The basic result for combining mean squares is that if --

$$n_i x_i / \sigma_i^2, \; i = 1,2,\ldots,k$$

(6, p.369) are independent chi-squared variables with n_i degrees of freedom, then --

$$\mu = ng/\tau$$

is approximately chi-squared with n degrees of freedom, where --

$$\tau = \Sigma g_i \sigma_i^2,$$

$$g = \Sigma g_i x_i, \text{ and}$$

$$n = (\Sigma g_i \sigma_i^2)^2 / \Sigma (g_i^2 \sigma_i^4 / n_i)$$

In applying this, the σ_i^2 are replaced by their estimators in the formula for n.

NOTATION FOR SINGLE-SITE CASES OF SECTION B1

Symbols subscripted by "a" or "p" have the meanings defined below, for the approved (old) or proposed (new) method, respectively.

Probability Distributions

$t_{df,\alpha}$ = the 100(α)th percentile of a t distribution with df degrees of freedom

$\chi^2_{df,\alpha}$ = the 100(α)th percentile of a chi-square distribution with df degrees of freedom

$F_{df_1,df_2,\alpha}$ = the 100(α)th percentile of an F distribution with df, numerator, df_2 denominator degrees of freedom.

Population Parameters

σ^2 = variance = (standard deviation)^2

μ = mean

b = bias = difference between the mean (expected value) of the measurement system under fixed conditions and the true value.

σ_o, b_o = data quality objectives, i.e., desired bounds on the standard deviation and bias of the proposed method.

Sample Statistics for One-Way ANOVAS

Y^1_j , Y^2_j = duplicate observations (e.g., recoveries or log measurements) on jth day for a particular concentration in the absolute case or method in the comparative case

D = number of days

N = total number of observations

$\overline{Y}_{ij}$, S_{ij} = the average and standard deviation of the (usually, two) observations in the (i,j) cell

MSW, MSB = within-day and between-day mean squares for one-way ANOVA designs with groups defined by different days

$s^2_{TOT} = g$ = estimate of total variance of a single observation under the model at hand (referred to in some formulae as "g" for compactness); for duplicate observations in each cell s^2_{TOT} is the simple average of MSB and MSW.

REFERENCES

1. Standard Practice for Dealing with Outlying Observations. ANSI/ASTM E178-80.

2. Box GEP, Hunter WG, Hunter JS. Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building. Wiley, New York. 1978.

3. Winer, B.J. Statistical Principles in Experimental Design. McGraw-Hill, New York. 1962.

4. Kleinbaum, D. G. and Kupper, L. L. Applied Regression Analysis. Duxbury Dress, North Scituate, Mass. 1978.

5. Draper, N. R. and Smith, H. Applied Regression Analysis. John Wiley and Sons, New York. 1966.

6. Graybill, F.A. An Introduction Linear Statistical Models. McGraw-Hill, New York. 1961.

APPENDIX C

EXCERPT FROM SW-846:

SECTION TEN, QUALITY CONTROL/QUALITY ASSURANCE

APPENDIX C

SECTION TEN

QUALITY CONTROL/QUALITY ASSURANCE

Section 10.1 defines Quality Control (QC) and Quality Assurance (QA). Section 10.2 discusses how QC/QA procedures can be used to ensure achievement of program goals. The various QC/QA aspects of sampling are discussed in Section 10.1.3 while Section 10.1.4 discusses and lists appropriate laboratory QC/QA activities. Section 10.1.5 discusses the criteria with which acceptable data must comply and methods of data evaluation.

10.1 Introduction

Quality assurance (QA) is a system for ensuring that all information, data, and resulting decisions compiled under a specific task are technically sound, statistically valid, and properly documented. Quality control is the mechanism through which quality assurance achieves its goals. Quality control programs define the frequency and methods of checks, audits, and reviews necessary to identify problems and dictate corrective action, thus verifying product quality.

The soundness of an organization's QC/QA program has a direct bearing on the integrity of its sampling and laboratory work. Results of sampling or analysis conducted without adequate quality control and assurance may be deemed unacceptable for RCRA evaluation purposes. The following section discusses some minimum standards for QC/QA programs. Generators who are choosing contractors to perform sampling or analytical work should make their choice only after evaluating the contractor's QC/QA program against the procedures presented in these sections. Likewise, contractors that currently sample and/or analyze solid wastes should similarly evaluate their QC/QA programs.

10.2 Program Design

The initial step for any sampling or analytical work should be to strictly define the program goals. Once the goals have been defined, a program must be designed that will meet these program goals. QC and QA measures will be the mechanisms used to monitor the program and to ensure that all data generated are suitable for their intended use. A knowledgeable person who is not directly involved in the sampling or analysis must be assigned the responsibility of ensuring that the QC/QA measures are properly employed.

As a minimum, a proper QC/QA program would include the following:

1. The intended use(s) for the data, and the necessary level of precision and accuracy of the data for these intended uses.

2. A representative sampling plan that includes provisions for:

   - selecting appropriate sampling locations, depths, etc.
   - providing a statistically sufficient number of sampling sites.
   - measuring all necessary ancillary data.
   - determining climatic flow or other conditions under which sampling should be conducted.
   - determining which media are to be sampled (e.g., wastewater, sediment, effluent, soil).
   - determining which parameters are to be measured (and where).
   - selecting appropriate sample containers.
   - selecting the frequency of sampling and length of sampling period.
   - selecting the types of sample (e.g., composites vs. grabs) to be collected.
   - sample preservation.
   - chain-of-custody.

3. An analytical plan that includes:

   - chain-of-custody procedures.
   - appropriate sample preparation methods.
   - appropriate analytical methods.
   - appropriate calibration and analytical procedures.
   - data handling, review and reporting.

4. Planning for the inclusion of proper and sufficient QC/QA activities, including the use of QC samples throughout all phases of the study to ensure that the level of quality of the data will meet the requirements of the intended use(s) of the data.

All program details should be put in writing and assignments made to appropriate personnel.

If the above procedures are followed (i.e., an appropriate program is designed, tasks are assigned to knowledgeable personnel, and sufficient QC/QA

steps are employed), the program should meet and possibly surpass its goals in most cases; at worst the failure to meet the program goals will be detected and the usefulness of any data will be quantified.

10.3 Sampling

The quality of a sampling program has a direct bearing on the legal, physical, and chemical integrity of the samples. If the representativeness of the samples cannot be verified due to inadequate attention to sampling procedures, then the usefulness of the analytical data will be limited, regardless of the refinement of the analytical program. It is imperative, therefore, that no analytical program be conducted without an adequate sampling plan which does or will document the degree of representativeness of the parameters of interest.

10.3.1 Design of a Sampling Plan

Section One of this manual discusses the considerations involved in designing a representative sampling plan. For each specific project, a sampling plan should be designed prior to commencement of sampling. If the plan addresses the considerations discussed in Section One, then the resulting samples should be representative of the waste of interest and therefore suitable for evaluation of the waste according to RCRA criteria.

10.3.2 Sample Collection

A variety of different sampling devices are used in sampling depending on the type of sample (solid, liquid, multiphased), the type of sample container, and the sampling location. Section One and portions of Section Three of this manual describe different devices that are available. The appropriate sampling device must be selected and its use supervised by a person thoroughly familiar with both the sampling and analytical requirements. This familiarity is essential since (1) certain sampling devices are made of materials that may contaminate samples, (2) cross contamination of samples can occur if the sampling device is not cleaned properly, (3) routine sampling methods may not be applicable when the waste is to be analyzed for a different parameter (e.g., volatile organic compounds), and (3) the method of employing the sampling devices may affect the integrity of the sample.

10.3.3 Sample Preservation

Some form of preservation is usually required for all samples. The type of sample preservation required will vary depending on the sample type and the parameter to be measured. Therefore, more than one container of the same waste may be necessary if the waste is to be analyzed for more than one parameter type.

The analytical methods included in this manual refer to the optimum means of preservation. Since the chemical make-up of certain samples can alter the effectiveness of preservation measures, all sample analyses should be performed as soon as possible after sampling and before any recommended holding time has expired.

10.3.4 Chain of Custody

Although chain-of-custody procedures may not be required in all cases, they often are an essential part of sampling/analytical schemes since these procedures can document the history of samples. Chain of custody establishes the documentation and control necessary to identify and trace a sample from sample collection to final analysis. Such documentation includes labeling to prevent mix-up, container seals to detect unauthorized tampering with contents of the sample containers, secure custody, and the necessary records to support potential litigation.

A sample is considered to be under a person's custody if (1) it is in the person's physical possession, (2) in view of the person, (3) secured by that person so that no one can tamper with the sample, or (4) secured by that person in an area that is restricted to authorized personnel.

Refer to Section One for details of how to implement chain-of-custody procedures.

10.4 Analysis

An analytical program defines standard operating procedures to be used in waste analysis, appropriate QC/QA procedures, means for detecting out-of-control situations, and remedial actions. A separate analytical program should be developed for each different waste to be analyzed. The program should be thoroughly specified before sampling is begun, since the analytical procedures to be used may affect the choice of sampling devices and procedures.

The program should select methods that will provide data at the level of accuracy and precision that will be required by users of the data for decision-making purposes under RCRA. Once the appropriate method(s) have been selected it is imperative that the accuracy and precision of all analytical data be thoroughly documented by means of a well-designed QC/QA program.

Laboratory QC/QA activities normally include:

1. Use of EPA-acceptable sample preparation and analytical methods.

2. Calibration of laboratory instruments to within acceptable limits according to EPA or manufacturer's specifications before, after, and during (as acceptable) use. Reference standards must be used when necessary.

3. Periodic inspection, maintenance, and servicing (as necessary) of all laboratory instruments and equipment.

4. The use of reference standards and QC samples (e.g., checks, spikes, laboratory blanks, duplicates, splits) as necessary to determine the accuracy and precision of procedures, instruments, and operators.

5. The use of adequate statistical procedures (e.g., QC charts) to monitor the precision and accuracy of the data and to establish acceptable limits.

6. A continuous review of results to identify and correct problems within the measurement system (e.g., instrumentation problems, inadequate operator training, inaccurate measurement methodologies).

7. Documenting the performance of systems and operators.

8. Regular participation in external laboratory evaluations (including the EPA Performance Audit Program) to determine the accuracy and overall performance of the laboratory. This should include performance evaluation and interlaboratory comparison studies, and formal field unit/laboratory evaluations and inspections.

9. Use of acceptable sample identification and, as necessary, formal chain-of-custody procedures in the laboratory.

10. Maintenance and storage of complete records, charts, and logs of all pertinent laboratory calibration, analytical, and QC activities and data.

11. Ensuring all data outputs are presented in their prescribed format.

Specific Quality Control measures for each method can be found by referring to the individual analytical methods included in this manual.

10.5 Data Handling

The quality of all data must be assessed before the data are used. Assessment should focus on five basic points.

1. Accuracy - Can the data's accuracy be determined, and is it acceptable for the planned use? QC/QA procedures will be designed to measure the accuracy of all analytical data.

2. Precision - Can the data's precision be determined, and is it acceptable for the planned use? QC/QA should demonstrate the reproducibility of the measurement process.

3. Completeness - Are a sufficient amount of data available for the planned use? QC/QA shall identify the quantity of data needed to meet the program goals.

4. Representativeness - How well do the data represent actual conditions at the sampling location, considering the original study design, sampling methods, analytical methods, etc., which were used?

5. Comparability - How comparable are data with respect to several factors, including:

   - consistency of reporting units?
   - standardized siting, sampling, and methods of analysis?
   - standardized data format?

All these factors must be considered when designing a study, and QC/QA procedures must specify a reviewing process for all data.

Statistical procedures applicable to data evaluation include:

1. Central tendency and dispersion

   - Arithmetic mean
   - Range
   - Standard deviation
   - Relative standard deviation
   - Pooled standard deviation
   - Geometric mean

2. Measures of variability

   - Accuracy
   - Bias
   - Precision; within laboratory and between laboratories

3. Significance test

   - u-test
   - t-test
   - F-test
   - Chi-square test

Specific data handling precautions are noted in the individual methods described in this manual.

☆U.S. GOVERNMENT PRINTING OFFICE: 1982-361-082/318

APPENDIX D

EXAMPLE TEST METHOD EQUIVALENCY PETITION

APPENDIX D

EXAMPLE TEST METHOD EQUIVALENCY PETITION

The following example represents a test method recently approved by EPA. The original report submitted to EPA[1] has been modified and supplemented to meet guidelines on petition format and content that were not available when it was first submitted. Fictitious information has been added in some areas and is identified by footnotes; other minor changes have been made for clarity and for conformity to guidelines.

This example demonstrates the format and general issues appropriate for an equivalency petition. It does not include an experimental design or statistical analyses of test data that the petitioner must submit. These analyses are often unique to the petition; therefore, analyses are excluded so as not to mislead the petitioner. Section 4 and Appendix B provide the guidance needed to establish experimental designs and statistical analysis procedures.

---

[1] A Laboratory Comparison Between EPA Method 450.1 [also referred to as Method 9020] and Haloscan TM for the Determination of Total Organic Halogen in Groundwater, Surface Water, and Waste Treatment Plant Effluents. Prepared by Richard A. Cope, Ph.D., for Environmental Research Group, Inc. Ann Arbor, Michigan. 1982.

OSWER Policy Directive
#9433.00-2

TEST METHOD EQUIVALENCY PETITION

TO MEASURE

TOTAL ORGANIC HALIDES (TOX) BY NEUTRON ACTIVIATION ANALYSIS

Petition Submitted By --

ENVIRONMENTAL RESEARCH GROUP, INC.

117 North First Street

Ann Arbor, Michigan 48104

[Petitioner should provide name and telephone number of a contact]

[Date]

INTRODUCTION

There is currently one existing published procedure for the determination of organic halogens (TOX) in water samples: EPA Method 9020. Any time an analytical method becomes accepted as an approved method, it is desirable to provide an alternative or additional analytical method, which is also approved, that can be substituted as equivalent to the first method. This is especially true if the first method specifies a specific brand of instrumentation such as the Dohrmann DX 20, which is specified in Method 9020.

This petition presents an alternative method to 9020 for the analysis of total organic halogens (TOX) using neutron activiation analysis. ERG believes the data show that this method produces equivalent data to Method 9020 while providing additional information on the organic halogen distribution of the sample. ERG requests that this method, called Haloscan™, be evaluated by the EPA and, if found adequate, be accepted as an approved procedure for the screening of groundwater, surface water, and waste treatment plant effluent for TOX.

1.0 CERTIFICATION OF ACCURACY AND RESPONSIBILITY

I certify under penalty of law that I have personally examined and am familiar with the information submitted in this demonstration and all attached documents and that based on my inquiry of those individuals immediately responsible for obtaining the information, I believe that the submitted information is true, accurate, and complete. I am aware that there are significant penalties for submitting false information including the possibility of fine and imprisonment.

Signed,

_____

Title _____

Date _____

2.0 DESCRIPTION OF PROPOSED ACTION

2.1 TEST METHOD DESCRIPTION

The proposed Haloscan™ test method is presented below in the format found in EPA's "Test Methods for Evaluating Solid Waste" (SW-846).

PROPOSED METHOD

TOTAL ORGANIC HALIDES (TOX) BY NEUTRON ACTIVATION ANALYSIS

1.0 Scope and Application

1.1 This method determines Total Organic Halides (TOX) in drinking waters, groundwaters, and river waters, and in wastewater treatment plant effluents. The method uses a carbon adsorption procedure identical to that of Method 9020 (TOX analysis using a microcoulometric-titration detector), irradiation by neutron bombardment, and then detection using a gamma ray detector. The reliable limits of detection are 5 ppb for chlorine and 1 ppb for iodine and bromine.

1.2 This method detects all organic halides containing chlorine, bromine, iodine, and fluorine that are adsorbed by granular activated carbon under the conditions of the method. Each halogen can be quantitated independently.

1.3 This method is restricted to use by, or under the supervision of, analysts experienced in the operation of neutron activation analysis and familiar with spectral interferences.

1.4 This method may be used in place of Method 9020 and has the advantage of determining the individual concentrations of the halogens chlorine, bromine, and iodine in addition to TOX.

2.0 Summary of Method

2.1 A sample of water that has been protected against the loss of volatiles by the elimination of headspace in the sampling container, and that is free of undissolved solids, is passed through a column containing 40 mg of granulated activated carbon (GAC). The column is washed to remove any trapped inorganic halides. The GAC sample is exposed to thermal neutron bombardment creating a radioactive isotope. Gamma ray emission, which is unique to each halogen, is counted. The area of the resulting peaks is directly proportional to the concentration of the halogens.

3.0 Interferences

3.1 Method interferences may be caused by contaminants, reagents, glassware, and other sample processing hardware. All these materials must be routinely demonstrated to be free from interferences under the conditions of the analysis by running method blanks.

3.1.1 Glassware must be scrupulously cleaned. Clean all glassware as soon as possible after use by treating with cromate cleaning solution. This should be followed by detergent washing in hot water. Rinse with tap water and distilled water, drain dry, and heat in a muffle furnace at 400° C for 15 to 30 min. Volumetric ware should not be heated in a

muffle furnace. Glassware should be sealed and stored in a clean environment after drying and cooling to prevent any accumulation of dust or other contaminants.

3.1.2 The use of high-purity reagents and gases helps to minimize interference problems.

3.2 Purity of the activated carbon must be verified before use. Only carbon samples that register less than 2,000 ng/40 mg should be used. The stock of activated carbon should be stored in its granular form in a glass container with a Teflon seal. Exposure to the air must be minimized, especially during and after milling and sieving the activated carbon. No more than a 2-week supply should be prepared in advance. Protect carbon at all times from all sources of halogenated organic vapors. Store prepared carbon and packed columns in glass containers with Teflon seals.

3.3 It is possible that other radioisotopes, stray radiation, counting geometries, and counting-equipment materials can affect gamma counting. It is essential that the data interpretation be performed by an analyst experienced at detecting these interferences.

4.0 Apparatus and Materials

4.1 Adsorption system

4.1.1 Dohrmann adsorption module (AD-2), or equivalent, pressurized, sample and nitrate-wash reservoirs.

4.1.2 Adsorption columns: Pyrex, 5-cm-long x 6-mm-O.D. x 2-mm-I.D.

4.1.3 Granular activated carbon (GAC): Filtrasorb-400, Calgon-APC or equivalent, ground or milled, and screened to a 100/200 mesh range. Upon combustion of 40 mg of GAC, the apparent-halide background should be 2,000 ng Cl^--equivalent or less.

4.1.4 Cerafelt (available from Johns-Manville), or equivalent: Form this material into plugs using a 2-mm-I.D. stainless-steel borer with ejection rod (available from Dohrmann) to hold 40 mg of GAC in the adsorption columns. CAUTION: Do not touch this material with your fingers.

4.1.5 Column holders (available from Dohrmann).

4.1.6 Volumetric flasks: 100-ml, 50-ml. A general schematic of the adsorption system is shown in Figure 1.

4.2 Containers suitable for containment of samples and standards during irradiation (e.g., 1/5-dr polyethylene snap-cap vial).

4.3 Sample introduction system and a reactor generating a thermal neutron flux capable of achieving enough halogen activity for counting purposes (e.g., a reactor having a neutron flux of 5 x 10^{12} neutrons/cm^2/sec and a pneumatic-tube sample introduction system).

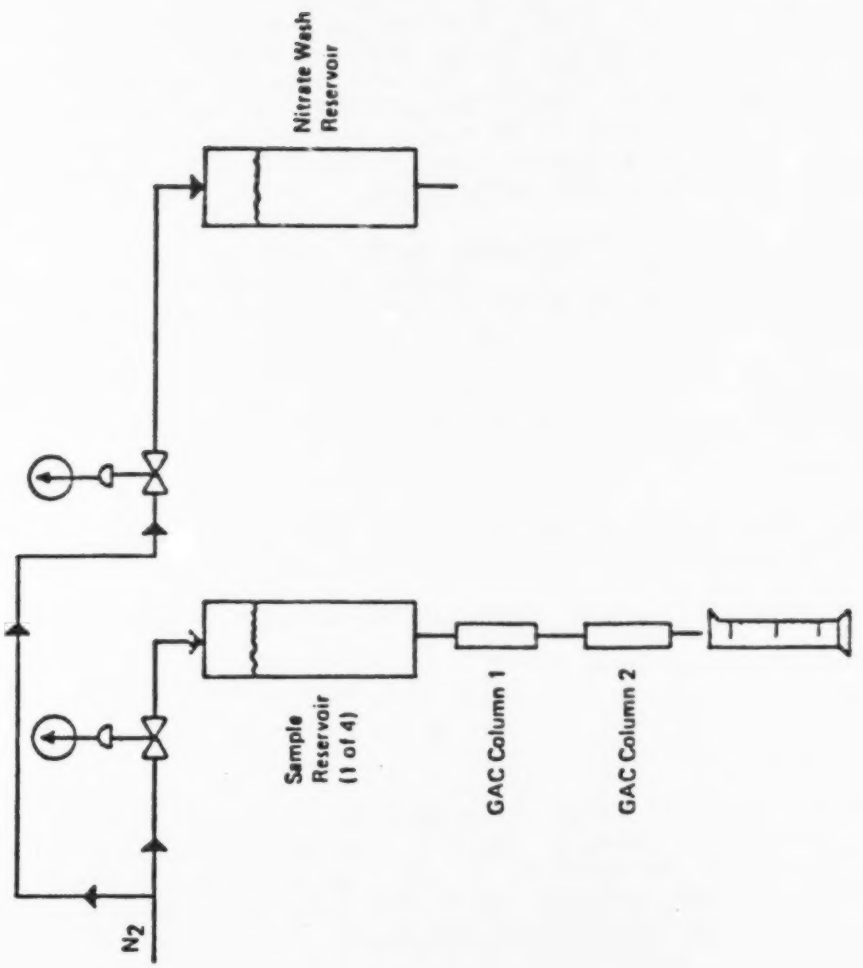


Figure 1. Schematic of absorption system.

4.4 A gamma ray detector and data handling system capable of resolving the halogen peaks from potential interferences and background.

5.0 Reagents

5.1 Prepurified nitrogen.

5.2 Deionized (DI) water.

5.3 500 ppm NO_3 solution (KNO_3 dissolved in DI water).

5.4 50 percent v/v mixture acetone and nanograde hexane.

5.5 0.1 M sodium sulfite (ACS reagent grade - 12.6 g/l).

5.6 Nitric acid: concentrated, reagent grade.

5.7 25-µg Cl, 2.5-µg Br, and 2.5-µg I standards.

5.8 Radioactive standards to be used for calibrating gamma ray detection systems.

6.0 Sample Collection, Preservation, and Handling

6.1 All samples must have been collected using a sampling plan that addresses the considerations discussed in Section One of SW-846.

6.2 All samples should be collected in bottles with Teflon septa (e.g., Pierce #12722 or equivalent) and be protected from light. If this is not possible, use amber glass, 250-ml, fitted with Teflon-lined caps. Foil may be substituted for Teflon if the sample is not corrosive. Samples must be protected against loss of volatiles by eliminating headspace in the container. If amber bottles are not available, protect samples from light. The container must be washed and muffled at 400° C before use to minimize contamination.

6.3 All glassware must be dried prior to use according to the method discussed in 3.1.1.

7.0 Procedure

7.1 Sample preparation

7.1.1 Special care should be taken in handling the sample in order to minimize the loss of volatile organohalides. The adsorption procedure should be performed simultaneously on the front and back columns.

7.1.2 Reduce residual chlorine by adding sulfite (1 ml of 0.1 M per liter of sample). Sulfite should be added at the time of sampling if the analysis is meant to determine the TOX concentration at the time of sampling. It should be recognized that TOX may increase on storage of the sample. Samples should be stored at 4° C without headspace.

7.1.3 Adjust the pH of the sample to approximately 2 with concentrated HNO_3 just prior to adding the sample to the reservoir.

7.2 Calibration

7.2.1 Check the adsorption efficiency of each newly prepared batch of carbon by analyzing 100 ml of the adsorption-efficiency standard, in duplicate, along with duplicates of the blank standard. The net recovery should be within 5 percent of the standard value.

7.2.2 Nitrate-wash blanks (method blanks): Establish the repeatability of the method background each day by first analyzing several nitrate-wash blanks. Monitor this background by spacing nitrate-wash blanks between each group of eight analysis determinations. The nitrate-wash blank values are obtained on single columns packed with 40 mg of activated carbon. Wash with the nitrate solution as instructed for sample analysis, and then analyze the carbon.

7.2.3 Prior to each day's operation, the instrument is calibrated using radioactive standards (e.g., cobalt-60 and radium-226 sources). The instrument is calibrated such that gamma rays from the standards fall within plus or minus one channel of their true energies. A 100-sec blank is then counted to verify that no stray radioactive sources are within sensing distance of the detector. As data are obtained throughout the day, peak locations in the standards are monitored to ensure that there is no electronic drift of the instrument. If drift is noted, the system is recalibrated.

7.3 Adsorption procedure

7.3.1 Connect two columns in series, each containing 40 mg of 100/200-mesh activated carbon.

7.3.2 Fill the sample reservoir, and pass a metered amount of sample through the activated-carbon columns at a rate of approximately 3 ml/min. NOTE: 100 ml of sample is the preferred volume for concentrations of TOX between 5 and 500 µg/l; 50 ml for 501 to 1,000 µg/l, and 25 ml for 1,001 to 2,000 µg/l.

7.3.3 Wash the columns-in-series with at least 2 ml of the 5,000-mg/l nitrate solution at a rate of approximately 2 ml/min to displace inorganic chloride ions.

7.4 Activation

7.4.1 After the quartz collection tube with the GAC is removed from the extraction unit, the GAC and cerafelt pads are extruded using the packing rod into a prewashed plastic container (e.g., 1/5-dr polyethylene snap-cap vial). The vial is prewashed to remove inorganic and organic chlorine by soaking in distilled water followed by storage in a glass jar containing 50 percent v/v acetone and hexane. Just prior to extrusion

the vial is removed by forceps and air-dried to remove residual water, acetone, and hexane. The vial is snapped shut, the hinge removed with a scalpel blade, the cap heat-sealed to the vial with an electric soldering gun reserved for that purpose, and a single-digit number placed on the vial with a Marks-A-Lot (or equivalent) marker pen.

7.4.2 Samples plus a similar vial containing 25 µg Cl, 2.5 µg Br, and 2.5 µg I standards are then introduced into the reactor, generally by placing them together in a 5-dr polyethylene vial and inserting them into a pneumatic-tube transfer "rabbit" for neutron irradiation. Irradiation would be for a 15-min period at a thermal neutron flux of 5×10^{12} neutrons/cm^2/sec. After returning from the reactor, the rabbit is allowed to "cool" for 20 min to allow short-lived radioisotopes (primarily Al) present in the GAC to decay away.

7.5 Detection

7.5.1 Analysis is performed using a lithium drifted germanium Ge(Li) gamma ray detector with an amplifier and a 4096-channel memory unit for data storage. The analyses can be performed either manually, with the operator changing samples and transferring the data to magnetic tape, or automatically, with both functions performed by an automatic sample changer.

7.5.2 Analysis begins by counting the standard and samples for a suitable time period (e.g., 200 sec "live" time for the standards and samples). The operator records the time intervals between samples and the "dead" time of each sample in a logbook for later use in calculating halogen concentrations in each sample.

7.6 Calculations

7.6.1 Chlorine, bromine, and iodine can be analyzed within a 200-sec counting period taking place 20 to 40 min after irradiation.

7.6.2 Chlorine is analyzed using the 1642-KeV gamma ray produced by 37.1-min ^{38}Cl. Bromine is analyzed using the 616-KeV gamma ray from 17.7-min ^{80}Br, while iodine is analyzed using the 442-KeV gamma ray produced by 25-min ^{128}I.

7.6.3 The calculation used for quantitation is:

$$\text{ppm halogen} = \frac{\text{cts unk.}}{\text{cts std.}} \times \frac{\text{counting time std.}}{\text{counting time unk.}} \times \frac{\mu\text{g in std.}}{\text{sample vol.}} \times e^{\lambda t}$$

where

cts unk. = the integrated area of the appropriate gamma ray peak in the unknown with background subtracted and the total multiplied by 1 + [(percent dead time unknown - percent dead time std.)/200]. The latter correction is usually less than 4 percent and corrects for pile-up errors.

cts std. = the integrated area of the appropriate gamma ray peak in the standard with background subtracted.

counting time std. = the "live" counting time in seconds of the standard.

counting time unk. = the "live" counting time in seconds of the unknown.

µg in std. = the number of micrograms of the stable element in question in the standard (25 for Cl, 2.5 for Br and I).

sample vol. = the volume of sample passed through the GAC column, in milliliters.

$e^{\lambda t}$ = the decay correction to bring all statistics back to t = 0. $\lambda = 0.693/t_{1/2}$ where $t_{1/2}$ = the half-life, in minutes.

t = the time interval in minutes from the end of the count of the standard until the end of the count of the sample.

7.6.4 No further calculations are necessary as long as the final sample is counted within 40 min after the end of irradiation. If samples are counted after 40 min, the addition to the 616-KeV peak of ^{80}Br from the 619-KeV peak from ^{82}Br becomes large enough that a correction factor must be applied. In practice, all counting should be completed in less than 40 min after irradiation.

8.0 Quality Control

(Refer to Section 3.4 of this example petition).

9.0 References

2.2 DESCRIPTION OF APPLICABLE SAMPLE OR MATRICES[1]

ERG is requesting that the proposed test method be approved to screen groundwater, surface water, and wastewater treatment plant effluent for organic halogens. In all cases, these three categories represent aqueous liquids. Only waste treatment plant effluent may be deemed an actual waste. Groundwater and surface water are merely analyzed for hazardous contaminants.

The groundwater analyzed for this petition was well water collected from aquifers in the State of Michigan. Raw well water (unchlorinated and unsoftened) was collected from aquifers at approximately 100- to 150-foot depths. These waters were generally hard and relatively high in iron.

---

[1] All information in Section 2.2 is fictitious.

The surface waters were collected from rivers in Michigan. These were hard waters, quite clear in appearance, with only minor suspended solids.

The waste treatment plant effluent samples were provided with the permission of municipal treatment plants in Michigan (all tertiary systems) from their effluent discharge into the rivers mentioned above.

[NOTE: The number of samples taken from each type of water and the number of locations sampled is determined during pre-petition negotiations between EPA and the petitioner.]

2.3 ASSESSMENT OF LIMITING OR INTERFERING FACTORS

Quality control of the extraction unit (i.e., the Dohrmann adsorption module (AD-2)) proved to be relatively straightforward with three exceptions: variation in sample size passed through the granular activated carbon (GAC) columns, the potential for channeling in the GAC columns, and the variation between GAC background levels from bottle to bottle and column to column because of difficulty obtaining the same volume of GAC in each column.

In general, the automatic shutoff feature of the AD-2 module worked satisfactorily, but sample would continue to pass slowly through the GAC column after shutoff due to gravity if the GAC columns were not removed immediately. This problem could be accounted for if the excess volume remained within the volumetric flask and could be measured, but if the flask overflowed the sample had to be discarded. To prevent overflow, it is recommended that only 1 or 2 milliliters of sample in excess of that to be analyzed be placed in the sample storage tube.

Two GAC tubes in sequence should always be used but they both also need to be analyzed if any detectable halogen shows up in the upper tube. We found occasionally (5 percent occurrence) that channeling or some other phenomenon would take place in the upper tube such that breakthrough to the lower tube would take place even though the GAC in the upper tube was not yet saturated with organic halogen.

The only area creating a potential problem was in the instrument stabilization step after insertion of the sample into the boat but prior to initiation of the analysis step. For good data reproducibility, it was found necessary to allow the MC-1 to stabilize to plus or minus two-digit variation in the TOX Det mode. This often took 5 to 10 minutes with the instrument appearing to plateau at a level (often about 3.5 on the panel meter) and then suddenly change to a lower value about 1.1 where pertinent stabilization would take place. For samples very high in organic halogens (>20 µg ultimate meter reading), stabilization was very slow and may be due to a continuous off gassing of small amounts of volatile organic halogens even prior to introducing the sample into the furnace.

For the actual neutron activation analysis, the amount of radioactivity induced in a sample is directly proportional to the neutron flux it is exposed to. Since this flux can vary depending on how the sample is positioned in relation to the reactor core during irradiation, it is essential that a known standard be irradiated with every sample batch to act as a flux monitor. This was done throughout the study. Since the sample acts as a point source emitting radiation (much as a light bulb viewed at a distance), the intensity of the radiation falls off with the square of the distance from the source. Thus, care must be taken to ensure that the standard and all samples associated with the standard are counted at the same distance from the detector. This is accomplished by using an automatic sample changer that can position the samples to within plus or minus 1 millimeter of the same position each time.

2.4 TEST METHOD QUALITY CONTROL (QC) PROCEDURES

2.4.1 All QC data should be maintained and available for easy reference or inspection.

2.4.2 Before performing any analyses, the analyst must demonstrate the ability to generate acceptable accuracy and precision with this procedure by analyzing appropriate QC check samples.

2.4.3 The laboratory must develop and maintain a statement of method accuracy for their laboratory. The laboratory should update the accuracy statement regularly as new recovery measurements are made.

2.4.4 Employ a minimum of one blank per sample batch to determine if contamination is occurring.

2.4.5 Run check standard after approximately every 15 samples.

2.4.6 Run 1 duplicate sample for every 10 samples. A duplicate sample is a sample brought through the whole sample preparation process.

2.4.7 It is recommended that the laboratory adopt additional quality assurance (QA) practices for use with this method. The specific practices that would be most productive will depend upon the needs of the laboratory and the nature of the samples. Field duplicates may be analyzed to monitor the precision of the sampling technique. Whenever possible, the laboratory should perform analysis of standard reference materials and participate in relevant performance-evaluation studies.

2.4.8 Quality control for the analysis phase is very straightforward since the instrument is a noncontact analyzer. That is, only the radiation emitted from the sample - not the sample - should touch the analyzer. Since contamination of the system is not usually a problem (unless a sample spills on it), the most serious QC issues deal with uniform neutron flux, counting geometry, and spectral interpretation. The amount of radioactivity induced in a sample is directly proportional to the neutron flux it is exposed to. Since

this flux can vary depending on how the sample is positioned in relation to the reactor core during irradiation, it is essential that a known standard be irradiated with every sample batch to act as a flux monitor. Care must also be taken to ensure that the standard and all samples associated with the standard are counted at the same distance from the detector.

3.0 STATEMENT OF NEED AND JUSTIFICATION FOR THE PROPOSED METHOD[1]

3.1 NEED

As discussed in the Statement of Interest, only one method for determining the presence of organic halogens is currently approved by EPA. This method (9020) is limited to one brand of instrumentation, the Dohrmann DX 20. We wish to provide an equivalent alternative to this analytical device, which can be substituted if necessary. This alternative is neutron activation analysis used via the Haloscan™ method.

3.2 JUSTIFICATION

This section compares the two analytical methods (the currently approved Dohrmann TOX method (9020) and the ERG Haloscan™ method for the analysis of organically bound halogens (chlorine, bromine, and iodine)) in three types of water samples: groundwater, surface water, and wastewater. The purpose of this exercise was to determine that the two methods produce essentially the same total halogen result and to verify that the Haloscan™ method provides additional information in that each halogen species is reported out independently of the others.

In the following pages, ERG presents comparability data showing the results of these two test methods. Precision, accuracy, and QC data are also provided. Sampling and test data on each replicate are provided in Section 3.3.

[NOTE: At this point, the petitioner should list the data quality objectives set by EPA and describe the experimental design established for this test program, addressing the number and timing of samples, locations, replicates, and testing, among other things.]

---

[1]The discussion as well as the data in this section is hypothetical.

Table 3-1A contains the test results from ground water source number 1 (Aquifer A)[1]. In each sample, the ambient levels of chlorine, bromine, and iodine were above the detection limit of the Dohrmann method. Therefore, it was not necessary to spike samples.

Tables 3-2A and 3-2B present the QC data collected during the analyses of ground water samples.[2] QC checks were made as outlined in Sections 7.4 through 7.7 of the test method description.

[NOTE: Here, the petitioner should present detailed information on statistical procedures used to evaluate test data. Also, raw data tables or computer printouts of the statistical analyses should be provided. Section 4 and Appendix B provide detailed information about this aspect of the petition.]

---

1 Table 3-1A contains fictitious data for one source of groundwater. In the interest of conserving paper, similar tables were not prepared for the remaining sources of ground water and the sources for both surface water and wastewater effluent. Thus, an actual petition, using the test design in this petition would have several tables that are similar to the format of Table 3-1A.

2 A complete petition would contain similar tables for surface water and wastewater effluent samples.

TABLE 3-1A. RESULTS OF GROUNDWATER ANALYSES:
SOURCE NO. 1, AQUIFER A (EXAMPLE TABLE)[1]

Samples 1-1, 1-2, 1-3: analyzed by ERG--1/1/85
Samples 1-4, 1-5, 1-6: analyzed by ABC Associates--1/1/85

Sample 1-1
(µg/100 ml)

Method 9020		Haloscan™ (Specific halogens)		
		Cl^-	Br^-	I^-
Replicate-1	1.53			
R-2	2.26			
Mean =	1.89			
Std. dev. =	0.52			
Replicate-3		1.86	<0.1	<0.1
R-4		2.30	<0.1	<0.1
Mean =		2.08		
Std. dev. =		0.31		

Sample 1-2
(µg/100 ml)

Method 9020		Haloscan™ (Specific halogens)		
		Cl^-	Br^-	I^-
Replicate-1				
R-2				
Mean =				
Std. dev. =				
Replicate-3				
R-4				
Mean =				
Std. dev. =				

Sample 1-3
(µg/100 ml)

Method 9020		Haloscan™ (Specific halogens)		
		Cl^-	Br^-	I^-
Replicate-1				
R-2				
Mean =				
Std. dev. =				
Replicate-3				
R-4				
Mean =				
Std. dev. =				

(Continued)

[1] The number of samples taken will vary based on prepetition negotiations. All data are fictitious.

TABLE 3-1A. RESULTS OF GROUNDWATER ANALYSES
SOURCE NO. 1, AQUIFER A (EXAMPLE TABLE)[1] (CONT.)

| Sample 1-4 |

(µg/100ml)

Method 9020		Haloscan™ (Specific halogens)		
		Cl^-	Br^-	I^-
Replicate-1	Replicate-3			
R-2	R-4			
Mean =		Mean =		
Std. dev. =		Std. dev. =		

| Sample 1-5 |

(µg/100ml)

Method 9020		Haloscan™ (Specific halogens)		
		Cl^-	Br^-	I^-
Replicate-1	Replicate-3			
R-2	R-4			
Mean =		Mean =		
Std. dev. =		Std. dev. =		

| Sample 1-6 |

(µg/100ml)

Method 9020		Haloscan™ (Specific halogens)		
		Cl^-	Br^-	I^-
Replicate-1	Replicate-3			
R-2	R-4			
Mean =		Mean =		
Std. dev. =		Std. dev. =		

[1] Data are fictitious.

TABLE 3-2A. QC DATA--GROUNDWATER ANALYSIS VIA METHOD 9020
(EXAMPLE TABLE)
(µg/100ml)

Samples	Blank (1 per sample)	Field Duplicate (1 per sample)	Lab Duplicate (1 per 10)	Standard (1 per 15)
Aquifer A				
1-1	____	____	N. A.	N.A.
1-2	____	____	____ (R-)	____ (R-)
1-3	____	____	____ (R-)	N.A.
1-4	____	____	____ (R-)	____ (R-)
1-5	____	____	____ (R-)	N.A.
1-6	____	____	N.A.	____ (R-)
Aquifer B				
2-1	____	____	____ (R-)	N.A.
2-2	____	____	____ (R-)	____ (R-)
2-3	____	____	____ (R-)	N.A.
2-4	____	____	____ (R-)	____ (R-)
2-5	____	____	N.A.	N.A.
2-6	____	____	____ (R-)	____ (R-)
Aquifer C				
3-1	____	____	____ (R-)	N.A.
3-2	____	____	____ (R-)	____ (R-)
3-3	____	____	____ (R-)	____ (R-)
3-4	____	____	N.A.	N.A.
3-5	____	____	____ (R-)	____ (R-)
3-6	____	____	____ (R-)	N.A.

N.A. = Not Applicable.

(Continued)

TABLE 3-2A. QC DATA--GROUNDWATER ANALYSIS VIA METHOD 9020 (CONT.)
(EXAMPLE TABLE)
(µg/100ml)

Samples	Blank (1 per sample)	Field Duplicate (1 per sample)	Lab Duplicate (1 per 10 replicates)	Standard (1 per 15 replicates)
Aquifer D				
4-1	____	____	(R-)	(R-)
4-2	____	____	(R-)	N.A.
4-3	____	____	N.A.	(R-)
4-4	____	____	(R-)	N.A.
4-5	____	____	(R-)	(R-)
4-6	____	____	(R-)	N.A.
Aquifer E				
5-1	____	____	(R-)	(R-)
5-2	____	____	N.A.	N.A.
5-3	____	____	(R-)	(R-)
5-4	____	____	(R-)	N.A.
5-5	____	____	(R-)	(R-)
5-6	____	____	(R-)	(R-)

N.A. = Not Applicable.

TABLE 3-2B. QC DATA--GROUNDWATER ANALYSIS VIA HALOSCAN™ METHOD
(EXAMPLE TABLE)
(µg/100ml)

Samples	Blank (1 per sample)	Field Duplicate (1 per sample)	Lab Duplicate (1 per 10 replicates)	Standard (1 per 15 replicates)
Aquifer A				
1-1	_____	_____	N.A.	N.A.
1-2	_____	_____	_____ (R-)	_____ (R-)
1-3	_____	_____	_____ (R-)	N. A.
1-4	_____	_____	_____ (R-)	_____ (R-)
1-5	_____	_____	_____ (R-)	N.A.
1-6	_____	_____	N.A.	_____ (R-)
Aquifer B				
2-1	_____	_____	_____ (R-)	N.A.
2-2	_____	_____	_____ (R-)	_____ (R-)
2-3	_____	_____	_____ (R-)	N. A.
2-4	_____	_____	_____ (R-)	_____ (R-)
2-5	_____	_____	N.A.	N.A.
2-6	_____	_____	_____ (R-)	_____ (R-)
Aquifer C				
3-1	_____	_____	_____ (R-)	N.A.
3-2	_____	_____	_____ (R-)	_____ (R-)
3-3	_____	_____	_____ (R-)	_____ (R-)
3-4	_____	_____	N.A.	N.A.
3-5	_____	_____	_____ (R-)	_____ (R-)
3-6	_____	_____	_____ (R-)	N.A.

N.A. = Not Applicable

(Continued)

TABLE 3-2B. QC DATA--GROUNDWATER ANALYSIS VIA HALOSCAN™ METHOD (CONT.)
(EXAMPLE TABLE)
(µg/100ml)

Samples	Blank (1 per sample)	Field Duplicate (1 per sample)	Lab Duplicate (1 per 10 replicates)	Standard (1 per 15 replicates)
Aquifer D				
4-1	____	____	(R-)	(R-)
4-2	____	____	(R-)	N.A.
4-3	____	____	N.A.	(R-)
4-4	____	____	(R-)	N.A.
4-5	____	____	(R-)	(R-)
4-6	____	____	(R-)	N.A.
Aquifer E				
5-1	____	____	(R-)	(R-)
5-2	____	____	N.A.	N.A.
5-3	____	____	(R-)	(R-)
5-4	____	____	(R-)	N.A.
5-5	____	____	(R-)	(R-)
5-6	____	____	(R-)	(R-)

N.A. = Not Applicable.